

STIC-Biotech/ChemLib

From: Rao, Manjunath N.
Sent: Tuesday, July 09, 2002 3:41 PM
To: STIC-Biotech/ChemLib
Subject: Request for sequence search for 09/784,554

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10C 01
Phone: 306-5681

Date: 7-9-02

Please search the following as soon as possible for application with serial number **09/784,554**

SEQ ID NO: 1, 3, and 5 against all commercial nucleic acid databases including issued patents database and pending application database and provide **a print of all results.**

SEQ ID NO: 2, 4, and 6 against all commercial protein databases including issued patents database and pending application database and provide **a print of all results.**

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: BOB
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 7-16-02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

1. The first part of the report deals with the general situation of the country and the progress of the work during the year.

2. The second part of the report deals with the results of the work during the year and the progress of the work during the year.

3. The third part of the report deals with the results of the work during the year and the progress of the work during the year.

4. The fourth part of the report deals with the results of the work during the year and the progress of the work during the year.

5. The fifth part of the report deals with the results of the work during the year and the progress of the work during the year.

6. The sixth part of the report deals with the results of the work during the year and the progress of the work during the year.

7. The seventh part of the report deals with the results of the work during the year and the progress of the work during the year.

8. The eighth part of the report deals with the results of the work during the year and the progress of the work during the year.

9. The ninth part of the report deals with the results of the work during the year and the progress of the work during the year.

10. The tenth part of the report deals with the results of the work during the year and the progress of the work during the year.

11. The eleventh part of the report deals with the results of the work during the year and the progress of the work during the year.

12. The twelfth part of the report deals with the results of the work during the year and the progress of the work during the year.

13. The thirteenth part of the report deals with the results of the work during the year and the progress of the work during the year.

14. The fourteenth part of the report deals with the results of the work during the year and the progress of the work during the year.

15. The fifteenth part of the report deals with the results of the work during the year and the progress of the work during the year.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 19:12:29 ; Search time 1513.98 seconds
(without alignments)
2427.980 Million cell updates/sec

Title: US-09-784-554B-5
Perfect score: 2141
Sequence: 1 atgaagcgcaaaataatagtag.....tttgttcccttagtgaggg 2141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2032	94.9	2141	22 AAH75061
2	1721.4	80.4	4056	22 AAH75060
3	1653.4	77.2	4059	22 AAH75059
4	571.4	26.7	1854	12 AAQ14842
5	570	26.6	2977	12 AAQ13001
6	336.6	15.7	6415	20 AAX55662
7	336.6	15.7	6416	24 AAD26526
8	61	2.8	3306	21 AAZ38862
9	60.6	2.8	3747	21 AAA67145

C 10	60.2	2.8	2643	19 AAV71207	DNA encoding ester
C 11	60	2.8	488	22 AAL01732	Human reproductive
C 12	60	2.8	547	22 AAD06925	5' end of human me
C 13	60	2.8	622	21 AAC81735	Human secreted pro
C 14	60	2.8	657	21 AAC79877	Human secreted pro
C 15	60	2.8	673	22 AAD06931	5' end of human me
C 16	60	2.8	678	22 AAK64492	Human immune/haema
C 17	60	2.8	711	21 AAC99831	Human secreted pro
C 18	60	2.8	836	22 AAL01738	Human reproductive
C 19	60	2.8	980	22 AAH32524	Human secreted pro
C 20	60	2.8	1301	19 AAV50430	Mouse coxsackievir
C 21	60	2.8	1397	21 AA261758	cDNA encoding rat
C 22	60	2.8	1397	22 AAC99691	Skin cell cDNA, SE
C 23	60	2.8	1442	22 AAK55538	Human immune/haema
C 24	60	2.8	1447	22 AAD13379	Human secreted pro
C 25	60	2.8	2335	22 AAI65505	Nucleotide sequenc
C 26	60	2.8	2452	22 AAK64491	Human immune/haema
C 27	60	2.8	2730	20 AAX84696	Human metastatic m
C 28	60	2.8	3124	22 AAF44826	RPP5-like protein
C 29	60	2.8	4517	22 AAH24427	PBKCMV phagemid
C 30	60	2.8	4518	19 AAV33626	Plasmid pBK-CMV DN
C 31	60	2.8	4752	22 AAH42272	Nucleotide sequenc
C 32	59.6	2.8	683	22 AAK64693	Human immune/haema
C 33	59.2	2.8	2109	22 AAD02809	HBXDJ03 cDNA clone
C 34	59	2.8	445	22 AAK55929	Human immune/haema
C 35	57.4	2.7	562	19 AAV23873	plant OMT enzyme D
C 36	57.4	2.7	562	20 AAZ06876	Pine O-methyl tran
C 37	57.4	2.7	562	21 AAA69980	Pinus radiata O-me
C 38	57.4	2.7	562	21 AAK67960	Pinus radiata OMT
C 39	56.8	2.7	859	21 AAC79869	Human secreted pro
C 40	55.8	2.6	616	22 AAD06927	5' end of human me
C 41	54.6	2.6	558	21 AAC98570	Human colon cancer
C 42	53.2	2.5	1775	12 AAQ13003	Endo3 gene encodin
C 43	52.8	2.5	1464	21 AAZ45341	DNA encoding a Hum
C 44	52.6	2.5	618	20 AAV84461	Human secreted pro
C 45	52.6	2.5	618	22 ABA83244	Human secreted pro

ALIGNMENTS

RESULT 1
AAH75061
ID AAH75061 standard; DNA; 2141 BP.
XX
XX
AC AAH75061;
XX
DT 29-OCT-2001 (first entry)
XX
DE Nucleotide sequence of xyloglucanase enzyme.
XX
KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring; ss.
XX
OS Paenibacillus pabuli.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..2085
FT
FT /product= "xyloglucanase"
FT /transl_except= "{pos: 715..822, aa: Xaa}"
FT /transl_except= "{pos: 835..837, aa: Yaa}"
FT /note= "no termination codon given; Xaa represents
36 unspecified residues, comprising amino
acids 239-274; Yaa represents an unspecified
residue"

WO200162903-A1.

30-AUG-2001.

21-FEB-2001; 2001WO-DK00116.

24-FEB-2000; 2000DK-0000291.
 (NOVO) NOVOZYMES AS.
 Schnorr K, Jorgensen PL, Schuelein M;
 WPT: 2001-522819/57.
 P-PSDB; AAG63964.
 New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 useful for detergent compositions, and textile or cellulose fiber
 processing industries
 Claim 14; Page 90; 97pp; English.
 The present sequence encodes a xyloglucanase of the invention. The
 specification describes a xyloglucanase enzyme belonging to family 44
 of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 of at least 30% at pH 5-8. The enzyme exhibits high performance in
 detergent compositions and prevents binding of certain soils to the
 xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 room temperature and has a half life of more than 50 days when incubated
 in a full formulated liquid detergent at 30 plus degrees celsius. The
 enzyme is used in detergent compositions, textile industry for improving
 the properties of cellulosic fibers, yarn, and woven or non-woven
 fabrics, preferably in textile scouring process, and in cellulose fiber
 processing industry for rattling of fibers e.g. hemp, jute, flax, and
 linen.
 Sequence 2141 BP; 566 A; 461 C; 531 G; 474 T; 109 other;

Query Match 94.9%; Score 2032; DB 22; Length 2141;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagcgaaataatagtagtaattttgtccaaacgttccaaatggtcgtcgtcgc 60
 Db 1 atgaagcgaaataatagtagtaattttgtccaaacgttccaaatggtcgtcgtcgc 60
 QY 61 atggatgcacgattatagtaggggggtctaccggctcccaactagttacacgtcaa 120
 Db 61 atggatgcacgattatagtaggggggtctaccggctcccaactagttacacgtcaa 120
 QY 121 acggcaaaagaccgttaccattaaagtcgataccaaagtcgttaagcctattagtcct 180
 Db 121 acggcaaaagaccgttaccattaaagtcgataccaaagtcgttaagcctattagtcct 180
 QY 181 tatatacaggtacgaatcagagatttggcaggcgatgaaatctggctccagagactt 240
 Db 181 tatatacaggtacgaatcagagatttggcaggcgatgaaatctggctccagagactt 240
 QY 241 ggtggcaatcgatgacggatgatacaactgggaaaaataatgtcccaatcgaggagcgat 300
 Db 241 ggtggcaatcgatgacggatgatacaactgggaaaaataatgtcccaatcgaggagcgat 300
 QY 301 tggcagcaatctagcgataactttttatgcaacaaatggtggcctgacaaaaacgcaatgt 360
 Db 301 tggcagcaatctagcgataactttttatgcaacaaatggtggcctgacaaaaacgcaatgt 360
 QY 361 gaaaacgggagcagtagcagacttcttcatgataccaaatcgatgaagctggcgcttat 420
 Db 361 gaaaacgggagcagtagcagacttcttcatgataccaaatcgatgaagctggcgcttat 420
 QY 421 tctttagtcacgtccgcatgcccgttatgtggcgaagatggaacgggaagtgtgcag 480
 Db 421 tctttagtcacgtccgcatgcccgttatgtggcgaagatggaacgggaagtgtgcag 480
 QY 481 gaaagcaacaggtccttccgtcgttggaaatcgaggttggtaaatgcaaaaatgcgcgcg 540
 Db 481 gaaagcaacaggtccttccgtcgttggaaatcgaggttggtaaatgcaaaaatgcgcgcg 540
 QY 541 ttccaactacagctgatctgaatgacaaatcagggtatagtcggatgaaatctgcaatttt 600

Db 541 ttccaactacagctgatctgaatgacaaatcagggtatagtcggatgaaatctgcaatttt 600
 QY 601 tttagtgaataatgacggcgctgtcttcaaaaagcggtgtgaaaggtatagtcgctgcac 660
 Db 601 tttagtgaataatgacggcgctgtcttcaaaaagcggtgtgaaaggtatagtcgctgcac 660
 QY 661 aatgaacccgctctctgtgtgcatacgcacgtccgcattctatggtagaaaagtctnnnnn 720
 Db 661 aatgaacccgctctctgtgtgcatacgcacgtccgcattctatggtagaaaagtctnnnnn 720
 QY 721 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 780
 Db 721 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 780
 QY 781 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 840
 Db 781 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 840
 QY 841 ggaacttctgtaaaaggcaactatagctgtgtctggactattacctggatcacaatgcgc 900
 Db 841 ggaacttctgtaaaaggcaactatagctgtgtctggactattacctggatcacaatgcgc 900
 QY 901 ctcaactcgaagcgaaggaagagattgtgtgatgtattcgtatgtgcactgttatccc 960
 Db 901 ctcaactcgaagcgaaggaagagattgtgtgatgtattcgtatgtgcactgttatccc 960
 QY 961 gaagcgatggcgagggcgcacgaattacaatgaggtaggcaatgacgaacgaagaaga 1020
 Db 961 gaagcgatggcgagggcgcacgaattacaatgaggtaggcaatgacgaacgaagaaga 1020
 QY 1021 gccagaatgcaggcgctctgtactttgtggaccgcgaactacaaggagaatagctggatc 1080
 Db 1021 gccagaatgcaggcgctctgtactttgtggaccgcgaactacaaggagaatagctggatc 1080
 QY 1081 gctcaatggaacagcgacttctgtcttactgctcgtgattgaaagcgctggtggacaag 1140
 Db 1081 gctcaatggaacagcgacttctgtcttactgctcgtgattgaaagcgctggtggacaag 1140
 QY 1141 tattaccgggaacaaagctggtttgaccgagtagtagtcacgctggcggaataatgata 1200
 Db 1141 tattaccgggaacaaagctggtttgaccgagtagtagtcacgctggcggaataatgata 1200
 QY 1201 tccggcggtattgtctatgaccgattgtgtggcctcttggcgaacaaacgactttatatg 1260
 Db 1201 tccggcggtattgtctatgaccgattgtgtggcctcttggcgaacaaacgactttatatg 1260
 QY 1261 gcgaactattggaagttaaaggatggtgcgaacaaactacgttagcgcgtcttacaagctt 1320
 Db 1261 gcgaactattggaagttaaaggatggtgcgaacaaactacgttagcgcgtcttacaagctt 1320
 QY 1321 taccgcaattatgacggaaaaaacgctactttcgcgatatacagcgttaatgcgcgaacg 1380
 Db 1321 taccgcaattatgacggaaaaaacgctactttcgcgatatacagcgttaatgcgcgaacg 1380
 QY 1381 tcggatattgtaatagctcgtgtgcatgtcttcgttaacggtatgcatctcacaagaactg 1440
 Db 1381 tcggatattgtaatagctcgtgtgcatgtcttcgttaacggtatgcatctcacaagaactg 1440
 QY 1441 cactctatgcatgaataaaagcatggaacgacgttcacgttcacgttcacgttcacgttc 1500
 Db 1441 cactctatgcatgaataaaagcatggaacgacgttcacgttcacgttcacgttcacgttc 1500
 QY 1501 ggcgagacgacttacagttccggttaaaataagggtcttcgataaaaaatagctcgcaaat 1560
 Db 1501 ggcgagacgacttacagttccggttaaaataagggtcttcgataaaaaatagctcgcaaat 1560
 QY 1561 aagcgatgagcgcaatcagcaaatctcagcaaatctcagcaaatcttaccatacagttacc 1620
 Db 1561 aagcgatgagcgcaatcagcaaatctcagcaaatcttaccatacagttaccatacagttacc 1620
 QY 1621 ttgacggttatcatcgtgttgcactgcgacgaatgatacacctgtgacacctgtgaa 1680

Db 1621 ttgacggtttatcaatcggttgactgcgcgaataatgatacacactgtgcacactgtgga 1680
 QY 1681 gatcctgaagctttacgctgaggtgaggtgcgcgaataatgatacacactgtgga 1740
 Db 1681 gatcctgaagctttacgctgaggtgaggtgcgcgaataatgatacacactgtgga 1740
 QY 1741 gacgtctccagcgaggttggttgactgacgctgacgctgacgctgacgctgacgct 1800
 Db 1741 gacgtctccagcgaggttggttgactgacgctgacgctgacgctgacgctgacgct 1800
 QY 1801 ttgctgctgagctgacgctgacgctgacgctgacgctgacgctgacgctgacgct 1860
 Db 1801 ttgctgctgagctgacgctgacgctgacgctgacgctgacgctgacgctgacgct 1860
 QY 1861 ggcactcttactattataataaataacgcaaaacaaagggggaacgaggaatccaat 1920
 Db 1861 ggcactcttactattataataaataacgcaaaacaaagggggaacgaggaatccaat 1920
 QY 1921 gtctgaaagcggttcgcggaacgctgtagacgcccggatcgctatgaagcggaagat 1980
 Db 1921 gtctgaaagcggttcgcggaacgctgtagacgcccggatcgctatgaagcggaagat 1980
 QY 1981 ggcacgctgaaggaacattgtgaatccagcggaacgcttctcggctactggttat 2040
 Db 1981 ggcacgctgaaggaacattgtgaatccagcggaacgcttctcggctactggttat 2040
 QY 2041 gtaactaatccacaaatgcaggggattccctgacgagcagcagcagcagcagcagc 2100
 Db 2041 gtaactaatccacaaatgcaggggattccctgacgagcagcagcagcagcagcagc 2100
 QY 2101 gcagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2141
 Db 2101 gcagcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2141

RESULT 2

AAH75060
 ID AAH75060 standard; DNA; 4056 BP.

XX
 AC AAH75060;

XX
 DT 29-OCT-2001 (first entry)

XX
 DE Nucleotide sequence of xyloglucanase enzyme.

XX
 KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
 XX
 KW cellulosic fiber; textile scouring; ss.

XX
 OS Paenibacillus polymyxa.

XX
 FH Key Location/Qualifiers
 FT CDS 1. 4056

FT FT
 FT FT

FT FT
 FT FT

FT FT
 FT FT

XX
 PN WO200162903-A1.

XX
 PN 30-AUG-2001.

XX
 PD 21-FEB-2001; 2001WO-DK00116.

XX
 PD 24-FEB-2000; 2000DK-0000291.

XX
 PA (NOVO) NOVOZYMES AS.

XX
 PI Schnorr K, Jorgensen PL, Schuelein M;

XX
 PI WPI; 2001-522819/57.

XX
 DR P-PSDB; AAG63963.

XX
 DR New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 XX
 PT useful for detergent compositions, and textile or cellulose fiber

PT processing industries

XX Claim 14; Page 85-86; 97pp; English.

XX The present sequence encodes a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
 CC linen.

XX Sequence 4056 BP; 1161 A; 888 C; 1070 G; 935 T; 2 other;

Query Match 80.4%; Score 1721.4; DB 22; Length 4056;

Best Local Similarity 88.6%; Pred. No. 0;

Matches 1855; Conservative 2; Mismatches 217; Indels 20; Gaps 5;

QY 1 atgaagcgcaaaatagtagtaattttgtccaaacgttccaaatggtcgtcgtcgc 60

Db 1 atgaagcgcaaaatagtagtagtaattttgtccaaacgttccaaatggtcgtcgtcgc 60

QY 61 atgagcatgcagattagtagggggtctcccggtccaaactgtagttcacggtcaa 120

Db 61 atgagcatgcagattagtagggggtctcccggtccaaactgtagttcacggtcaa 120

QY 121 acggcaaaagaccgttaccattaaagtgcatacatccaaagatcgtaagcctattagct 180

Db 121 acggcaaaagaccgttaccattaaagtgcatacatccaaagatcgtaagcctattagct 180

QY 181 tatatacaggtacgaatcaggatttggcagcgatgaaatctggtcgcacagactt 240

Db 181 tatatacaggtacgaatcaggatttggcagcgatgaaatctggtcgcacagactt 240

QY 241 ggtggcaatcgatgaccgatacaactgggaaataataatgtccaaatgcggaagcgat 300

Db 241 ggtggcaatcgatgaccgatacaactgggaaataataatgtccaaatgcggaagcgat 300

QY 301 tggcagcaatctagcgataactttttatgcacaatgtggtcgtgcacaaagcgaatgt 360

Db 301 tggcagcaatctagcgataactttttatgcacaatgtggtcgtgcacaaagcgaatgt 360

QY 361 gaaaagcgggagcagtgacgacttctttcatgatcaatcgctgagctggcgcttat 420

Db 361 gaaaagcgggagcagtgacgacttctttcatgatcaatcgctgagctggcgcttat 420

QY 421 tctttagtcacgctgcccgttggttggtgcaagatggaacgaaagtgtgcag 480

Db 421 tctttagtcacgctgcccgttggttggtgcaagatggaacgaaagtgtgcag 480

QY 481 gaaagcgaacaggtcctcctccgctggtggaatcaggtcgtaaatccaaaaatgcgcg 540

Db 481 gaaagcgaacaggtcctcctccgctggtggaatcaggtcgtaaatccaaaaatgcgcg 540

QY 541 ttccaaactacagctgactgaatgacaatcaggtatgctgaggaattcgcaatttt 600

Db 541 ttccaaactacagctgactgaatgacaatcaggtatgctgaggaattcgcaatttt 600

QY 601 ttagtgaacaaagtcagcgctgcttccaaacaaagcggtgtgaaagatatgcgtcgac 660

Db 601 ttagtgaacaaagtcagcgctgcttccaaacaaagcggtgtgaaagatatgcgtcgac 660

QY 661 aatgaacccgctctctgttcgatacgcattccgcgcatcattcgttgtaaaagtc----- 714

Db 661 aatgaacccgctctctgttcgatacgcattccgcgcatcattcgttgtaaaagtc----- 714


```

Db 2034 tggttatgaacgaattcccaacccaggggattctctgacgatgacattca 2087
|||||
RESULT 4
AAQ14842
ID AAQ14842 standard; DNA; 1854 BP.
XX
AC AAQ14842;
XX
DT 11-FEB-1992 (first entry)
XX
DE B.lautus endol core domain/H.insolens 43kD endoglucanase CBD fusion.
XX
KW cellulose; carbohydrate binding domain; fungi; cellulase;
KW endoglucanase; EGI; fusion gene; linking B region; ss.
XX
OS Bacillus lautus.
OS Humicola insolens.
XX
FH Key Location/Qualifiers
FT CDS 1..1854
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT misc_feature /*tag= c
FT /*tag= d
FT /*note= "Ser 34 to Val 549 of B.lautus endol core"
FT /*note= "43kD B region and CBD"
XX
PN WO9117244-A.
XX
PD 14-NOV-1991.
XX
PF 08-MAY-1991; 91WO-DK00124.
XX
PR 09-MAY-1990; 90DK-0001158.
XX
PA (NOVO ) NOVO NORDISK A/S.
XX
PI Woldike HF, Hagen F, Hjort CM, Hastrup S;
XX
DR WPI; 1991-353766/48.
DR P-PSDB; AAR15241.
XX
PT New fungal (hemi) cellulose degrading enzymes - for prodn. of
PT liq. fuel gas and feed protein, have specified carbohydrate
PT binding domain
XX
PS Claim 11; Fig 15; 73pp; English.
XX
CC This fusion construct encodes the 43kD signal peptide and the first
CC 4 N-terminal amino acids of the mature ca. 43kD endoglucanase from
CC H.insolens fused to the core of Endo 1 produced by B.lautus NCIMB
CC 40250 followed by the peptide Ile-Ser-Glu (encoded by a linker)
CC fused to the 43kD linking B region and carbohydrate binding domain.
CC The fusion protein represents one example of a cellulose-degrading
CC enzyme of the invention. The recombinant DNA construct encoding it
CC is used to transform Aspergillus oryzae.
CC See AAQ14838-Q14841.
XX
SQ Sequence 1854 BP; 452 A; 509 C; 500 G; 393 T; 0 other;

Query Match 26.7%; Score 571.4; DB 12; Length 1854;
Best Local Similarity 59.4%; Pred. No. 1e-171;
Matches 917; Conservative 0; Mismatches 615; Indels 12; Gaps 3;

QY 133 gttaccattaaatcgatacatccaggatcgtaagcctattagtccttatatacgt 192
|||||

```

```

Db 82 gtcaatttcacgattataacgcactcggaacgtgcagcgatcagccccaattattcggga 141
QY 193 acgaatcagatttgcagcgcatgaaatctggtcgcagacgacttggtagcaatcga 252
|||
Db 142 accaatcaggtatcggcgagcgaggaactggtccctcccgaggtcggagcaaccgg 201
QY 253 atgaccgggatacaactgggaaaaataatattgtccaatgcgggaagcgattggcagcaatct 312
|||
Db 202 ctgacgggttacaaactgggagaaacgcattccagcgccggaaggaactggcttcattac 261
QY 313 agcgataaattttatgcacaatggtgacctgacaaaagccgaaatgtgaaaagccggga 372
|||
Db 262 agcgatgattctctgcggcaacggtgtgttccgacacccgactcgacaaacggcg 321
QY 373 gcagtgcagacttcgttcatgatcaatcgctgaagctggcgcttattcttagtcacg 432
|||
Db 322 gcggtgttaccgctttcacgataaaattttgcagaatggagcttactccattgtaacg 381
QY 433 ctgccgatggcgggttatgtgccaaggatggaacggaagtgtgcagaaagcgaaacg 492
|||
Db 382 ctgcaaatggcggttatgtgtccgggataaagaaacggtccagttgacgagagtgagacg 441
QY 493 gctccttcgcctcgttgggaatcaggtcgttaaatgccaataatgcgcgttccaactacag 552
|||
Db 442 gctccctcacccggttgggataaggtcgagtttgcacaaatgcgcgttctcccttca 501
QY 553 cctgatctgaatgacaatcaggttatatcggtatcggaattcgtcaatttttagtgaagaag 612
|||
Db 502 cctgatctgaacgagcaacagtgtatgtatgataagaagttaacttctcgttcaacgg 561
QY 613 taaggcgctgttcaacaaagcggtgtgaaagatatgctgcgcgaatgaacccgct 672
|||
Db 562 tatggaacgcttcaacgtcaacggcgcatcaaaagcgattctcggtgataacagcgcgcg 621
QY 673 ctctgtgcacatccgcgcattcatggtgaaaaagtcnntnnnnnnnnnnnnnnnnnn 732
|||
Db 622 ctgtggtcgtgagacgcatccaggattccatccggagcgagttacaaagcgagaaactc 681
QY 733 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 792
Db 682 gctaagagcatcgactgtcacaaggcgctgaagaacgctgcagtcgcagcgaataattc 741
QY 793 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 846
|||
Db 742 ggtctgcctttacgggttctcgcgcatattgtctctgcaggaacgacggattggcg 801
QY 847 tctgtaaaagcaactatagctgttctggtgactattacgtgatacaaatgcgcctcaac 906
|||
Db 802 agtttgaaggcaactacagctcggtttatcgattactatctggtacagatgaagaatgct 861
QY 907 tcgcaagccgaaggcaagagattgctggtatgtattcgatgtgcactggttatcccgagcg 966
|||
Db 862 catacgagaacgcaaaagattgctgattgctggagctccactggttatccggaagca 921
QY 967 atggcgaggagcagcagcaattacaaatga---ggtaggcaatgcagaaacgaagaagcc 1023
|||
Db 922 caggcgaggcgagcagcaatcgtcttggcgggcgcggaataatgcatacgcagaagct 981
QY 1024 aqaatgcagggcctcgtacttctgtggacccgacactacaaggaagatagctgactgct 1083
|||
Db 982 cgcgtacaaagcccaagatcgctatgggactcggcttccaccgggaagacagctgctgctgc 1041
QY 1084 caatggaacagcgcatcttctgcttactgctcgtatgaaagcagctggttggaacagtat 1143
|||
Db 1042 acatggttttcaagctacttgccttaattccgaagctgcaattcttcgattcagacgtat 1101
QY 1144 taccgggaacaaagctgtgtttgacagagtatagctacggcgcgaaaaatgattttcc 1203
|||
Db 1102 taccgggtacgaagctgctgacagatcagagttcagctacggcggaacaatcatttgc 1161
QY 1204 ggcgttatgtatgacccgatgtcgtggcatcttggcacaacaaacgacttatatggcg 1263
|||
Db 1162 ggagcgatagctaccgcgacgctcgctcgtcttcttggaaaaatattgctgttcgcgcg 1221

```


QY	193	acgaatcaggatttggccaggcgatgaaaaatctgctgccagacgacttggtygcaatcga	252
Db			
QY	4390	gcaaatcaggatctcagggtgttg--ttcacctctcaagacgacttggtyggaacaga	4446
Db			
QY	253	atgaccggtacaaactgggaaaaataatgtccaatcggaagcgacttgccagcaatct	312
Db			
QY	4447	ttagcgggttacaattggggaacaataatgtccaatcagggagtgactggttatctta	4506
Db			
QY	313	agcgataactttttatgcaacaatggtgcctgcacaaagccgaatgtgaaagccggga	372
Db			
QY	4507	agcgatgattatgtgttatattggtataacagggaatgataagaacggttccagca	4566
Db			
QY	373	gaagtgaacacttggttcattgatcaatctgctgaagcttggcgtttatttttagtcaag	432
Db			
QY	4567	gctgttgttaagcaaatcttcagcagcagtcacataaagcaaaatgcattcaggcatcaca	4626
Db			
QY	433	ctgcgcgatggcgggttatgttgccaaagatggaacggaagtgtgcaggaagcgcaacag	492
Db			
QY	4627	ttaacagatggtaggttatgttgccaaagatgggaatggttcagtgagcagtcagagaca	4686
Db			
QY	493	gtccttcgcgtcgttggaatcaggctgtaaatgccaaaaatgcgcgttcccaactacag	552
Db			
QY	4687	gtccgctgccgagatggcctgaggtcaagtttaaaaagatggtgcactgtcattgcag	4746
Db			
QY	553	ctgtacttgaaatgacaaatcaggatgatgcgatgaatttcgttcaatttttttagtgaaaaag	612
Db			
QY	4747	ctgcagctgaaatgaataactatgtatataatggtatggtttaaactatctgttaataag	4806
Db			
QY	613	tacggcgctgcttcacaaaagcggtgttgaaagatatgcgtccgacatgaaccgcct	672
Db			
QY	4807	tatggtcgatcatctgtgcacgggaattaaaggaatatacttgacacgagccggac	4866
Db			
QY	673	ctctggttcgatacgcattccgcgcatctcatggtgaaaaagtcnnnnnnnnnnnnnnnn	732
Db			
QY	4867	ttatggttactactactccgcgaattcaccagaaggtaaacctgcagtgatgata	4926
Db			
QY	733	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn	792
Db			
QY	4927	aataaatcgttgagctggcgaaagtataaaagacacttgatccagatgcagaaatttt	4986
Db			
QY	793	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn	846
Db			
QY	4987	ggacctgcacgtatggttttgggatatattaaactatgcaggatgcacctgactggaat	5046
Db			
QY	847	tctgtaaaaggcaactatagctggttcgtggaactattaccttgatcaaaatgcctccaac	906
Db			
QY	5047	caggtttaaaggaaatcacagatggtttttgagctggttaccttgagcagatgaagaagca	5106
Db			
QY	907	tcgcaagccgaagcgaagatgtctgagatgtattctgatgtgcacttggtatcccaagcg	966
Db			
QY	5107	tcggatgatttgggaaraggttatcttggaatgacttgcacatacacttggtaaccggaggcg	5166
Db			
QY	967	atggcgaggagcattacaaaatgaggtaggcaatgacgaacgaagaa-----	1019
Db			
QY	5167	caggttggcggtgtgcgaatatgtttgcggtggaataagtacttcaagggtatggcca	5226
Db			
QY	1020	--agccagaatgcaggcgctctacttttgggaaccgcacctacaag-----	1065
Db			
QY	5227	atagcaggtatgcagccagcagacgctatgggataccgcacataaaaccaccagaaa	5286
Db			
QY	1066	-----gaagatagctgagctcaatggaaacagcgattttgct	1107
Db			
QY	5287	ggtcagataaacagcgggagaaaaatgcttggaataaaccaatggtttccagagtatctcca	5346
Db			
QY	1108	ttactgcctcgtattgaagcagctcggtggaacaagtattaccgcgggaacaaagctggcttg	1167
Db			
QY	5347	ctgcttcccaataaaagccagatatagacaagattatcctggtaccacaaacttgctata	5406
Db			
QY	1168	accgagtatagtaacggcggaataatgatatattccggcggttatgtctatgacagatgtg	1227
Db			
QY	5407	actgagtttgatttgagggaaggaacacatatatcgggaggaataagcttttagcagatgtg	5466
Db			

QY	1228	ctggcgatcttgggcaaaaacgacgctttatatggcgaactatttggaagttaaaggatcggt	1287
DB	5467	ttagggatattcggaagttagtatcatatggcagcaagatg-----gggagat	5517
QY	1288	gccaaacaactagttagcgccgttacaaagctttaccgcgaattatgacggaaaaaacgct	1347
DB	5518	tcggggagctatgacagcgggcgctacaacattttatctcaactatgatggaaaggttcg	5577
QY	1348	acttcggcgatcatcagcgttaattcgcaaacgtcggatatgtttaatagctcggtgcat	1407
DB	5578	agatacggttcaacgtgttgagcgctgagacaactgacgttgagaacatgcgcggtatat	5637
QY	1408	gcttcgctaacggtatgcctctacaagaactgcaccttcattgtcatgaataaaagcatg	1467
DB	5638	gcttcaattgagggaagatgattcgactgtgcataattatttaattacaacgaattat	5697
QY	1468	gacagcgcatctgagcgcccaattcgatctttccggcgagacgacttacagttccggtaaa	1527
DB	5698	gacaggaaactgaaggcagagataaagatgaataaccagggttatacacacagtgaggag	5757
QY	1528	atatggggcttcgataaaaaatagctcgcaaatgaagcagtagcgccaatccacgcaaat	1587
DB	5758	atatacggatttgacagtacagctctcagatcagggaagtgggagtgctcagtaatat	5817
QY	1588	tcaggcaacgcgtttacctatcacagtacacacctttgacggcgttatcacatcgttgtact	1647
DB	5818	caaaacacacaatcaccatagaagttccaaatctgacggtataccatatgttttaact	5877
QY	1648	gc 1649	
DB	5878	tc 5879	
RESULT 8			
AAZ38862/c			
ID	AAZ38862 standard; DNA; 3306 BP.		
XX	AAZ38862;		
AC			
XX			
DT	17-FEB-2000 (first entry)		
XX			
DE	Human Jurkat cell clone P2-15 AIM10 which affects TRRE activity.		
XX			
KW	Human; Jurkat cell; tumour necrosis factor receptor releasing enzyme;		
TRRE; cytokine; TNF; identification; cytostatic; anti-inflammatory;			
KW	cardiant; immunomodulator; antiarthritic; antibacterial; cancer;		
KW	heart failure; cachexia; inflammation; endotoxic shock; arthritis;		
KW	multiple sclerosis; sepsis; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	W0958559-A2.		
XX			
PD	18-NOV-1999.		
XX			
XX	14-MAY-1999; 99W-US10793.		
XX			
XX	14-MAY-1998; 98US-0081385.		
PR			
XX			
PA	(REGC) UNIV CALIFORNIA.		
XX			
PT	Catanaga T, Granger GA;		
XX			
XX	WPI; 2000-039067/03.		
DR			
XX			
PT	Tumour necrosis factor receptor releasing enzyme modulators and		
PT	polynucleotides		
XX			
PS	Claim 2; Page 68-69; 106pp; English.		
XX			
CC	The present invention describes isolated polynucleotides (A) comprising		
CC	sequence expressed at the mRNA level in Jurkat T cells and showing		
CC	increased enzymatic activity for cleaving and releasing the tumour		

CC necrosis factor (TNF) receptor in genetically modified COS-1 cells
 CC expressing the receptor. The present sequence represents a specifically
 CC claimed clone which affects tumour necrosis factor receptor releasing
 CC enzyme (TRRE) activity. Methods from the present invention can be used
 CC to assess a disease condition associated with altered TRRE activity. The
 CC polypeptides, polynucleotides and antibodies can be used to decrease or
 CC increase signal transduction from a cytokine in a cell. The polypeptides,
 CC polynucleotides and antibodies may be used to treat heart failure,
 CC cachexia, inflammation, endotoxemic shock, arthritis, multiple sclerosis
 CC and sepsis, and cancer.
 XX
 SQ Sequence 3306 BP; 694 A; 1024 C; 1020 G; 567 T; 1 other;

Query Match 2.8%; Score 61; DB 21; Length 3306;
 Best Local Similarity 100.0%; Pred. No. 2.8e-08;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2081 cgatccactagtgtcgacctgcagcgcgagctccagcttttttcccttttagtgagg 2140
 |||||
 Db 61 CGATCCACTAGTGTGCGACCTGCAGCGCGGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGG 2
 QY 2141 g 2141
 |
 Db 1 G 1

RESULT 9
 AA67145/c
 ID AAA67145 standard; DNA; 3747 BP.
 AC AAA67145;
 DT 31-OCT-2000 (first entry)
 XX
 DE Eucalyptus grandis cellulose synthase nucleotide sequence SEQ ID NO:141.
 XX
 KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
 KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
 KW transgenic plant; ds.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200022092-A2.
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-NZ00169.
 XX
 PR 13-OCT-1998; 98US-0170862.
 PR 11-AUG-1999; 99US-0148426.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Bloksberg LN;
 XX
 DR WPI: 2000-339328/29.
 DR P-PSDB; AAB16338.
 XX
 PT New genes encoding proteins involved in a plant polysaccharide
 PT biosynthetic pathway, useful for modulating or altering the
 PT polysaccharide content, composition or structure of the plant
 XX
 PS Claim 1; Page 106-107; 301pp; English.
 XX

CC The present invention describes isolated polynucleotides (PN) comprising
 CC a sequence selected from one of 835 nucleotide sequences given in
 CC AA67073 to AA67907, their (reverse) complements, sequences producing
 CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
 CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
 CC 835 sequences or sequences that are degenerately equivalent or allelic
 CC to the 835 sequences. The polynucleotides are used to modify the

CC activity of a polypeptide involved in a polysaccharide biosynthetic
 CC pathway in the plant. They are especially used to modulate or alter the
 CC polysaccharide content, composition or structure of the plant. AAB16268
 CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
 CC given in the present invention.
 XX
 SQ Sequence 3747 BP; 976 A; 822 C; 935 G; 1014 T; 0 other;

Query Match 2.8%; Score 60.6; DB 21; Length 3747;
 Best Local Similarity 88.0%; Pred. No. 4.1e-08;
 Matches 66; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2067 ttccctgacgatgcacatccactagtgtcgacctgcagcgcgagctccagcttttgt 2126
 |||||
 Db 404 TCCCGCAATTCCTTTGGATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGT 345
 QY 2127 tcccttttagtgagg 2141
 |||||
 Db 344 TCCCTTTAGTGAGG 330

RESULT 10
 AA71207/c
 ID AA71207 standard; DNA; 2643 BP.
 XX
 AC AA71207;
 XX
 DT 08-FEB-1999 (first entry)
 XX
 DE DNA encoding esterase E005 from compost isolate C-1.
 XX
 KW Esterase; enzyme stability; specificity; biocatalyst;
 KW resolution method; ss.
 XX
 OS Unidentified.

XX
 FH Key Location/Qualifiers
 FT CDS 1064..2598
 FT /*tag= a
 FT /product= esterase
 XX
 PN WO9846770-A2.
 XX
 PD 22-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US07237.
 XX
 PR 11-APR-1997; 97US-0827810.
 XX
 PA (THER-) THERMOGEN INC.

XX
 PI Aikens J, Allen L, Casadaban M, Demirjian D, Fonstein M;
 PI Vonstein V;
 XX

DR WPI: 1998-594584/50.
 DR P-PSDB; AAW85078.
 XX

XX
 PT New isolated esterase encoding nucleic acids - obtained from
 PT microbial strains obtained from soil and compost samples, used for
 PT producing esterases for hydrolysis of chemical compounds
 XX
 PS Claim 6; Fig 6F; 103pp; English.
 XX

CC The present sequence encodes an enzyme having esterase activity.
 CC The nucleic acid sequence is obtained from an organism isolated
 CC from compost isolate C-1. The esterase has a range of substrate
 CC specificities (including regioselectivity, stereoselectivity), enhanced
 CC enzyme stability, and can be produced in large quantities for commercial
 CC use. The esterase can be used particularly in biocatalyst based
 CC resolution methods.
 CC note: the CDS feature given is that for the longest protein decoded.
 CC Other alternative start sites are indicated in the specification, with

AC AAD06931;
XX
DT
XX
DE 06-AUG-2001 (first entry)
XX
DE 5' end of human melanoma tumour-associated antigen cDNA from clone 5.31.
XX
KW Human; melanoma tumour-associated antigen; cytostatic; gene therapy;
KW vaccine; cancer; immune response; metastatic melanoma; immunotherapy; ss.
XX
OS Homo sapiens.
XX
PN WO200129064-A2.
XX
PD 26-APR-2001.
XX
PF 18-OCT-2000; 2000WO-US41260.
XX
PR 18-OCT-1999; 99US-0160042.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Strong TV, Conry RM, Lobuglio AF;
XX
DR WPI; 2001-308473/32.
XX
XX Isolated DNA sequence encoding a melanoma tumor associated antigen,
PT useful for detection, diagnosis and staging of melanomas, monitoring
PT metastatic melanomas and as a target for immunotherapy -
XX
PS Claim 1; Fig 2G; 55pp; English.
XX
XX The present sequence is 5' end cDNA of clone 5.31 encoding human melanoma
CC tumour-associated antigen. The cDNA is useful in gene therapy and as
CC an oligonucleotide probe for detecting mRNA coding for the melanoma
CC tumour-associated antigen in a sample. The cDNA is useful for
CC vaccinating an individual who is at risk of getting cancer, suspected of
CC having cancer or has cancer. The present sequence is useful for
CC inhibiting growth of melanoma tumour by inducing an immune response
CC against it. The melanoma tumour-associated antigen is useful for
CC detection, diagnosis and staging of melanoma, monitoring metastatic
CC melanoma, as new targets for immunotherapy. The identification of novel
CC tumour antigens may allow recurrence and metastatic disease to be
CC detected and disease burden monitored. Characterisation of the melanoma
CC tumour-associated antigen is of particular use in melanoma research.
XX
SQ Sequence 673 BP; 163 A; 201 C; 141 G; 158 T; 10 other;

Query Match 2.8%; Score 60; DB 22; Length 673;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2082 gatccactagtgtcagctcagcgagcgagctccagcttttttcccttttagtgagg 2141
DB 96 gatccactagtgtcagctcagcgagcgagctccagcttttttcccttttagtgagg 37

Search completed: July 12, 2002, 19:12:58
Job time: 31655 sec

NAME/KEY: CDS
LOCATION: 677..2776
OTHER INFORMATION:
US-07-862-588B-1

Query Match 26.6%; Score 570; DB 2; Length 2977;
Best Local Similarity 59.2%; Pred. No. 5.3e-165;
Matches 921; Conservative 0; Mismatches 624; Indels 12; Gaps 3;

Qy 120 aacggcaaacggtaccattaaagtcgatacatcaagagatcgtaagcctattagtc 179
Db 769 AGCTGCGAGTGATGTCATTTCAGATTAATACGAGTCGGAACGTCGACGATCAGCCC 828
Qy 180 ttatatatacgtacgaatcagatttggcagcgatgaaatctgctgccaagcact 239
Db 829 CAATATTATACGGAACCAATCAGGATCTGACGGGACGGAGAACTGGTCATCCCGCAGGCT 888
Qy 240 tgggtggcaatcgaatgaccgatacacaactgggaaaaataatgtcccaatgcygggaagcga 299
Db 889 CGGAGGCAACCGGCTGACGGGTTTACAACTGGGAGAACACGCATCCAGCGCGCGGAAGGGA 948
Qy 300 ttggcgaacatctagcgaatacctttttatgcaacaatggtggcctgacaaaaagccgaatg 359
Db 949 CTGGCTTCATTACAGCGATGATTTCTCTCGCGCAACGGTGGTTCACAGACACCGACTG 1008
Qy 360 tgaagccgggagcagtcgacgactctcttcgatgatacaatcgtaagctggcgctta 419
Db 1009 CGACAGCGCGGGCGGTTGTTACCGCTTTTACGATTAATCTTTGGAGAAATGGAGCTTA 1068
Qy 420 ttctttagtcacgctgcgagtcgagcggtgtatgtggccaaggtatggaaacggaagtgtgca 479
Db 1069 CTCCTATTGTAACGCTGCAAAATGGCGGTTATGTCTCCCGGATAAAGAACGCTCCAGTTGA 1128
Qy 480 ggaaggaacagcgtctcttcgctcgttggaatcaagtcgtaaatcccaaaatcgcd 539
Db 1129 CGAGAGTGAGACGGCTCCGTCACCGCTGGGTAAGTCGAGTTTGCCAAAATGCGCC 1188
Qy 540 gtcccaactacagctgctatcgtgaatgacaatcaggtatcggtatggaatctgcaattt 599
Db 1189 GTTCTCCCTTCAGCGCTGATCTGACGACGACGACGATGATATGATGAGAAAGTTAACTT 1248
Qy 600 tttagtgaaaaagtcagcgcgctgttccaaacaaagggcggtgtgaaaggtatagcgtcga 659
Db 1249 CTTGGTCAACCGGTATGGAACCGCTTCAACGCTCAACCGGGCATCAAGCGTATTTCGCTGGA 1308
Qy 660 caatgaacccgctctctggtcgatcagcgcgctcattcaggtgtaaaagtcnncnnn 719
Db 1309 TAACGAGCGCGGCTGTGGTCTGAGACGCTCAAGGATTCATCCGAGGAGCTTACAGC 1368
Qy 720 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 779
Db 1369 GGCAGAACTCTGCTAAGAGCATCGACTTGTCAAGCGGTGAAGAACGTCGATCGCGCA 1428
Qy 780 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 833
Db 1429 TGCCGAAATATTCGGTCTGCGCTTTACGGTTTCGGGCGCATATTGTCTCTGACGAGGACG 1488
Qy 834 cntgattggaactctgtlaaaggcaacta tagctggtgtggtgactattacctggatca 893
Db 1489 ACCGATTTGGCGGATTTGCAAGCACTACAGCTGTTTATCGATTACTATCTGGATCA 1548
Qy 894 aatgcgcctcaactcgaagccgaagcgaagatgtgctggatgtatcgtatgcaactg 953
Db 1549 GATGAAGAATGCTCATACGCAAGACGCAAAAGATTGCTGATGTGTGGACGCTCCACTG 1608
Qy 954 gtatcccaagcagtgagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1010
Db 1609 GTATCCGGAAGACAGGCGGAGGCGGAGCGAGCAATCGTCTTTGGCGGGCGGCGCAATATCGA 1668
Qy 1011 aacgaagaaagccagaa tgcagggcgctctgactttgtggaccccgacctacaggaaga 1070
Db 1669 TACGAGAGGCTCGCTGACAGCGCCCAAGATCGCTATGGGATCCGCTTACCAGGAAGA 1728

Qy 1071 tagctggatcgctcaatggaacagcgcatcttcttactgctgctgctgctgctgctgctgctgct 1130
Db 1729 CAGCTGGATCGGCACATGGTTTCAAGCTACTTGGCCCTTAATTCGAAGCTCGAATCTTC 1788
Qy 1131 ggtgacaagtattaccggggaacaaagctggcttggaccagagtagatagctcagcgggcga 1190
Db 1789 GATTTCAGACGTTATTCGGGGTACGAAGCTGGCGATACAGAGTTTCAGCTACGCGCGAGA 1848
Qy 1191 aaatgatatttcggcgatattgctatgacagatgctgagcatcttgggcaaaacga 1250
Db 1849 CAATCACATTTTCGGGAGGATAGTACCGCGGACCGCTCGGCATTTTGGAAATATGG 1908
Qy 1251 cgtttatatgggaaactattggaagttaaaagatggtgccaacaaactcgttagcgcgcg 1310
Db 1909 CGTTTATGCCGGAATTTACTGSCACAGGAGGACAATACCGATTATACC---AGCGCTGC 1965
Qy 1311 ttacaagctttaccgcaattatgacgggaaaaaacgctacttttcggcgatatacagcgtaa 1370
Db 1966 TTACAAAGCTGTATCGCAACTAGCAGCGCAATAAATCGGGGTTCGGCTCGATCAAAAGTGA 2025
Qy 1371 tgcgcaaacgctcgatattgttaatagctcggtgcatcttcctgtaacgagtgcatccta 1430
Db 2026 CGCCGCTAGCTCGATACGGAGAACAGCTCGTATACGCTTCGTTAACTGACGAGGAGA 2085
Qy 1431 caaagaactgcacctcattgtcatgaaataaagcattggaacgagcgcgcatcgagcccaatt 1490
Db 2086 TTCCGAACTCCACCTGATCGTCTGAATAAATAAATTTTCGACGATCGATCAACGCTACTTT 2145
Qy 1491 cgaatcttcggcgagagacattacagttccggttaaaatatgggcttcgataaaatag 1550
Db 2146 CCAGCTGTCTGTGTATAAACCTACACATCCCGGAGAGTATGGGCTTCGACCAACCGG 2205
Qy 1551 ctgcgaataaaggcagtagcgcccaatcacgcaaatcttcaggcaaacccgtttaccctatac 1610
Db 2206 ATCCGACATTCGGAACAGACGCTATACGAATATTAAACAACAAATCAATTCACGTATAC 2265
Qy 1611 agtacacatttgagcgcttatacatogtctgactgctgacacatgatacaactgt 1667
Db 2266 GCTTCTCCATTGTGGCTTACCACATTTCTGTGAAGCGGATAGCACCAGACCGGT 2322

RESULT 2

US-08-361-920-28
; Sequence 28, Application US/08361920
; Patent No. 5457046
; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 54570460 No. 5457046disk of No. 5457046th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
 TITLE OF INVENTION: Or Hemicellulose
 NUMBER OF SEQUENCES: 85
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.
 STREET: 405 Lexington Avenue, 62nd Floor
 CITY: New York
 STATE: New York
 COUNTRY: United States of America
 ZIP: 10174-6201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/479,939
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/361,920
 FILING DATE: 22-DEC-1994
 APPLICATION NUMBER: US/07/940,860
 FILING DATE: 28-OCT-1992
 APPLICATION NUMBER: DK 1158/90
 FILING DATE: 09-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/DK91/00124
 FILING DATE: 08-MAY-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 3435.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-867-0298
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1854 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Bacillus lautus/Humicola insolens
 FEATURE:

NAME/KEY: CDS
 LOCATION: 1..1851
 US-08-479-939-28

Query Match 26.8%; Score 569.8; DB 1; Length 1854;
 Best Local Similarity 59.3%; Pred. No. 4.7e-165;
 Matches 916; Conservative 0; Mismatches 616; Indels 12; Gaps 3;

QY 133 gtaccattaaatcgatatacatccaaggatcgtaaacgtatttagtcttatatacagtc 192
 DB 82 GTCACTTTCAGGATTAATACGAGTCGGAGACGTCGAGCGATCAGCCCAATATTACGGA 141
 QY 193 acgaatcaggattggcaggcgatgaaatcttgctgcccagacgacttggtggcaatcga 252
 DB 142 ACCAATCAGATCTGAGCGGAGGAGACTGGTCAATCCGCGGAGGCTCGGAGCAACCGG 201
 QY 253 atgacggatatacaactgggaaataatagtccaatgcgggagcgttggtggcaatct 312
 DB 202 CTGACGGGTTCACACTGGGAGAACACGATCCAGCGCGGAGGAGGACTGGCTTATTAC 261
 QY 313 accgataactttttatgcaacaatggtgacctgacaaaaagcgaatgtgaaaaagcgga 372
 DB 262 ACCGATGATTTCTGCGGCAACGGTGTGTTCAGACACCGACTGCCACAACGCGGGG 321
 QY 373 gcagtgacgactctgtttcatgatcaatcgctgaagctggcgcttattcttagtcagc 432

DB 322 GGGTGTATTACCGCTTTTACGATAAATCTTTGGAGATGGAGCTTACTCCATTGTACCG 381
 QY 433 ctgcgatgagccggttatgtgcccgaagatgcaaacgaagtgtgcaggaagcaaacag 492
 DB 382 CTGCAAAATGGCGGTTATGTGTCCCGGATAGAAGACGGTCCAGTTGACGAGAGTGAGAGC 441
 QY 493 gtccttccgctcgttgggaatcaggtcgtaaatgcaaaaaatgcgcgttcccaactacag 552
 DB 442 GCTCCGTCACCGGTTGGGATGAAGTTCGAGTTTGCACAAAATGCGCGTTCCTCCCTTCAG 501
 QY 553 cctgatctgaatgacaatcaggtatatgcggtgatgaattcgtaatttttagtgaataaag 612
 DB 502 CCGTGTCTGAACGACGAGCAAGTGTATGATGAAGAGTTAACTTCTGTGTCACACGG 561
 QY 613 tacggcgctgcttcaaaaaagcggtgtgaaagatatcgctctgcaactgaacccgct 672
 DB 562 TATGNAAGCTTCAACGTCACGCGCATCAAAAGGCTATTCGCTGGATAACGAGCGGG 621
 QY 673 ctctggtcgcatcacgcattccgcgcatcattggtgaaaagtcnntnnnnnnnnnnnnnn 732
 DB 622 CTGTGCTCTGAGACGATCCAAAGGATTCATCGGAGCAGTTACAAGCGGACAGACTCGTC 681
 QY 733 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 792
 DB 682 GCTAAGAGCATCGACTTGTCAAAAGCGGTGAAGAACGTCGATCCGCATGCCGAAATATTC 741
 QY 793 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 846
 DB 742 GGTCTGCGCTTTACGGTTTCGGCGCATATTTGCTCTGCAGGACGACCGGATTTGGCGG 801
 QY 847 tctgtaaaaggcaactatagctgtgtggaactattaccctggatcaaatgcgctcaac 906
 DB 802 AGTTTCAAGGCAACTACAGCTGTTTATCGATTACTACTGTGATCAGATCAAGAATGCT 861
 QY 907 tcgcaagcgaagcaagagattgctggtatgctgctgctgctgctgctgctgctgctgctg 966
 DB 862 CATACGACGACGCAAGAAAGTGTCTGATGTGCTGGAGCTCCACTGTGATTCGGGAAGCA 921
 QY 967 atggcgaggagcattacgaattacaaatga---ggtaggcaatgacgaacgaagaagcc 1023
 DB 922 CAGGCGGAGGCGACGGAATCGTCTTTCGGCGGCGGCAATATCGATACGACGANGCT 981
 QY 1024 agaatgcaggcgctctgacttctggtggaacccgacctacaaggaagatagctgctgctgct 1083
 DB 982 CGCGTACAAAGCGCAAGATCGCTATGGATCGCGCTTACAGGAAGACAGCTGATCGCG 1041
 QY 1084 caatggaacagcgacttctgcttctgctgctgctgctgctgctgctgctgctgctgctgct 1143
 DB 1042 ACATGTTTTCAAGCTACTTGGCTTAATTCGGAAGCTGCAATCTTCGATTCAGACGAT 1101
 QY 1144 taccggggaacaaagctgcttggaccgagtagctacggtggtggtggtggtggtggtggtggt 1203
 DB 1102 TATCGGGTACGAGCTGGCGATCACAGATTCAGTACGCGGAGAGCAATACATATTCG 1161
 QY 1204 ggcgtatgctatgacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1263
 DB 1162 GGAGCATAGTACGCGGAGCGCTCGGCTATTTGGAAAATATGGCGTTTATGCGCGG 1221
 QY 1264 aactattggaagttaaaagatgggtgcaaacactacgttagcgcttcaaacgcttcaac 1323
 DB 1222 AATTACTGGCAG---ACGAGGACAAATACCGATTATACCGCGCTGCTTCAAGCTGTAT 1278
 QY 1324 cgaattatgacggaacaaacgctacttctgctgctgctgctgctgctgctgctgctgctgctg 1383
 DB 1279 CGCAACTACGACGCAATTAATCGGGTTCGGCTCGATCAAGTGGAGCGCGCTACGTC 1338
 QY 1384 gatatgttaatagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1443
 DB 1339 GATACGGAAGACAGCTCGGTATAGCTTTCGGTAACTACAGGAGGAATTCGGAACCTCCAC 1398
 QY 1444 ctcatgtcatgaataaaagcatgacagcgacttcgacgcccattcgcattcttccttcggc 1503

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1624 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match 2.4%; Score 51.6; DB 2; Length 1624;
Best Local Similarity 52.3%; Pred. No. 7.2e-06;
Matches 114; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 1695 tacgctgagggtgagctggcgatgggaaagtcgatttgcctggagcgttccagcg 1754
DB 1058 TAACCTCTCGCAACAGCGGCGCAACGCCAGGTATCATTAACCTGGAACGAGTTAGCGG 1117
QY 1755 agttgtgggttacagtgtagcgggcaacgtatgaaacgggtcctttgtgctgtagc 1814
DB 1118 GCGCAGCAGCTATACCGTAAAGCGAGCAACGAGCGGCGTCCGTACACGAATGTGA 1177
QY 1815 atccaacttggtcgaacgttttacgatacgaacgtacgaacgcaacttcttacta 1874
DB 1178 CCGGGGTGTCACGCGCAGAGCTACAGCAACCGGGCTGACGAATGGCAGCAGTATTA 1237
QY 1875 ttataataaccgcaaaaacaaaggcggaacgagcg 1912
DB 1238 TTATGCTGAGGCAATCCAAATACCGGGCGGAGCAGCG 1275

RESULT 11
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796

GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1775 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..(1625.1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
US-07-862-588B-5

Query Match 2.4%; Score 51.6; DB 2; Length 1775;
Best Local Similarity 52.3%; Pred. No. 7.6e-06;
Matches 114; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 1695 tacgctgagggtgagctggcgatgggaaagtcgatttgcctggagcgttccagcg 1754
DB 1058 TAACCTCTCGCAACAGCGGCGCAACGCCAGGTATCATTAACCTGGAACGAGTTAGCGG 1117
QY 1755 agttgtgggttacagtgtagcgggcaacgtatgaaacgggtcctttgtgctgtagc 1814
DB 1118 GCGCAGCAGCTATACCGTAAAGCGAGCAACGAGCGGCGTCCGTACACGAATGTGA 1177
QY 1815 atccaacttggtcgaacgttttacgatacgaacgtacgaacgcaacttcttacta 1874
DB 1178 CCGGGGTGTCACGCGCAGCAGCTACAGCAACCGGGCTGACGAATGGCAGCAGTATTA 1237

Qy 1875 ttataaaataaccgcgaacaaacgaagcgagcg 1912
||||| + ||| + ||||| + |||||
Db 1238 ttatgctgtagggcatccaaatagcggcgagcgcg 1275

RESULT 12
 US-08-651-579-1/c
 : Sequence 1, Application US/08651579
 : Patent No. 5912141
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: BROJATSCH, JURGEN
 : APPLICANT: NAUGHTON, JOHN
 : APPLICANT: YOUNG, JOHN A. T.
 :
 : TITLE OF INVENTION: TUMOR VIRUS SUSCEPTIBILITY GENES AND GENE
 : PRODUCTS, AND USES RELATED THERETO
 :
 : NUMBER OF SEQUENCES: 2

Query Match	2.3%	Score 48.6;	DB 2;	Length 2413;
Best Local Similarity	65.6%	Pred. No. 7.5e-05;		
Matches	86;	Conservative 0;	Mismatches 44;	Indels 1; Gaps 1;
QY	2011	agcgggacgcgggtcttcggtactggttataactaatttcacaaatgcaggggattcc	2070	
Db	205	AGCGGAGCGCTGCGCCCGGCGTGCAGATCGGATTCGGAGCCAGTATAGGAACCTCGTGC	146	
QY	2071	ctgacgatgacgatccactagtgtcgcactgcagcgcgagctccagcttttgttccc	2130	
Db	145	CGAATTCCTTGGATCCACTAGTGTGCACCTGCAGGC-CGGCAGCTCCAGCTTTGTTC	87	
QY	2131	tttagtgaggg	2141	
Db	86	TTTAGTGAGGG	76	

RESULT 13
US-09-188-930-230/c
; Sequence 230, Application US/09188930A
: Patent No. 6150502

```

; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188.930A

```

Query Match 2.2%; Score 47.6; DB 3; Length 2004;
Best Local Similarity 96.7%; Pred. No. 0.00014;
Matches 58: Conservative 1; Mismatches 0; Indels 1; Gaps 1;

RESULT 14
US-09-400-208B-4/c
; Sequence 4, Application US/09400208B
; Patent No. 6271011
; GENERAL INFORMATION:
; APPLICANT: Lee, Margie
; APPLICANT: Sanchez, Susan
; APPLICANT: Henk, Adam
; TITLE OF INVENTION: Pasteurella Neureminidase Coding Sequences,
; TITLE OF INVENTION: Compositions and Diagnostic Methods
; FILE REFERENCE: 77-96B
; CURRENT APPLICATION NUMBER: US/09/400.208B

Query Match	2.1%	Score 45.2;	DB 4;	Length 2001;
Best Local Similarity	94.0%;	Pred. No. 0.00075;		
Matches 47: Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

RESULT 15
US-08-975-316-48/c

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 13:39:47 ; Search time 11269 Seconds
(without alignments)
2564.285 Million cell updates/sec

Title: US-09-784-554B-5
Perfect score: 2141
Sequence: 1 atgagcgcaaaatagtag.....tttctcccttagtgagg 2141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	60	2.8	250	10	BE418492
C 2	60	2.8	404	10	BE418409
C 3	60	2.8	473	10	BE910013
C 4	60	2.8	498	10	BE418489
C 5	60	2.8	567	10	BE418408
C 6	60	2.8	657	9	AT1815377
C 7	60	2.8	703	10	BE661741
C 8	60	2.8	832	10	BE418461
C 9	60	2.8	902	10	BE418314
C 10	60	2.8	934	10	BE417925
C 11	60	2.8	960	10	BE418436
C 12	60	2.8	999	10	BE418406
C 13	60	2.8	1027	10	BE418240
C 14	60	2.8	1043	10	BE418877
C 15	60	2.8	1058	10	BE418493
C 16	60	2.8	1114	10	BE418079
C 17	60	2.8	1152	10	BE418407

C 18	59	2.8	731	10	BE418460
C 19	58.4	2.7	679	10	BG907904
C 20	58.4	2.7	862	9	AF041408
C 21	58	2.7	703	10	BE660013
C 22	57	2.7	428	12	BH001150
C 23	55.8	2.6	730	10	BE418405
C 24	55.8	2.6	808	10	BE418018
C 25	54.8	2.6	638	10	BG908638
C 26	53.2	2.5	562	10	BE418186
C 27	49	2.3	721	10	BE418426
C 28	48.8	2.3	479	10	BE418339
C 29	48	2.2	288	10	BI468109
C 30	46.4	2.2	455	12	AZ133777
C 31	46.4	2.2	590	9	AI816608
C 32	45	2.1	508	9	AI816601
C 33	45	2.1	515	9	AI816613
C 34	45	2.1	537	9	AI816609
C 35	44.2	2.1	574	10	BG908735
C 36	44	2.1	526	9	AI816625
C 37	44	2.1	528	9	AI816614
C 38	44	2.1	560	9	AI816635
C 39	43.6	2.0	670	9	AI816672
C 40	43.4	2.0	521	9	AI816621
C 41	43	2.0	385	10	BE418396
C 42	43	2.0	528	9	AI816629
C 43	42.4	2.0	522	9	AI816636
C 44	42.4	2.0	694	9	AI816677
C 45	42.2	2.0	658	9	AI816673

ALIGNMENTS

RESULT 1
BE418492/c
LOCUS BE418492 250 bp mRNA linear EST 24-JUL-2000
DEFINITION SCL034.E08R990601 ITEC SCL Wheat Leaf Library Triticum aestivum
cDNA clone SCL034.E08, mRNA sequence.
ACCESSION BE418492
VERSION BE418492.1 GI:9416338
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE 1 (bases 1 to 250)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.K., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticaceae
JOURNAL Unpublished (2000)
COMMENT Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@agr.ca
International Triticaceae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .250
/organism="Triticum aestivum"
/cultivar="Thatcher Lrl"
/db_xref="taxon:4565"
/clone="SCL034.E08"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVRI"

```

BASE COUNT      55 a      84 c      76 g      34 t      1 others
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 250;
Best Local Similarity 100.0%; Pred. No. 6.1e-07;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2082 gatccactagtgtcgacctgcaggcgccgagctccagctttgtcccttttagtgagg 2141
|||||
Db 82 GATCCACTAGTGTGACCTGCGAGCGCGGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGG 23

RESULT 2
BE418409/c
LOCUS
DEFINITION SCL032.F11R90531 ITEC SCL Wheat Leaf Library Triticum aestivum
CDNA clone SCL032.F11, mRNA sequence.
ACCESSION BE418409
VERSION BE418409.1 GI:9416255
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 404)
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrrells, M., Warburton, M. and Wenzel, G.
TITLE International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
JOURNAL Unpublished (2000)
COMMENT Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .404
/organism="Triticum aestivum"
/cultivar="Thatcher Lrl"
/db_xref="taxon:4565"
/clone="SCL032.F11"
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT      89 a      117 c      93 g      77 t      28 others
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2082 gatccactagtgtcgacctgcaggcgccgagctccagctttgtcccttttagtgagg 2141
|||||
Db 83 GATCCACTAGTGTGACCTGCGAGCGCGGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGG 24

RESULT 3
BE910013/c
LOCUS
DEFINITION TaLr1124A09R TaLr1 Triticum aestivum cDNA clone TaLr1124A09 5',
mRNA sequence.

```

```

ACCESSION BG910013
VERSION BG910013.1 GI:14317689
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 473)
AUTHORS Cloutier, S., Dong, G. and Walsh, A.
TITLE Wheat functional genomics- Thatcher Lr1 cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >2.2 kb
Plate: 124 row: A column: 09
Seq primer: M13 Reverse.
Location/Qualifiers
1. .473
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="TaLr1124A09"
/clone_lib="TaLr1"
/tissue_type="Leaf tissue"
/dev_stage="14 Days old"
/lab_host="E. coli XL0LR"
/note="Vector: Lambda ZapII; mass excised in plasmid
vector pBK-CMV (Stratagene).; Site 1: EcoRI; Site 2: XhoI;
mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
inoculation with leaf rust pathogen Puccinia triticiina
race BBB carrying the avirulence gene Avr1."

BASE COUNT      111 a      153 c      146 g      63 t
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2082 gatccactagtgtcgacctgcaggcgccgagctccagctttgtcccttttagtgagg 2141
|||||
Db 67 GATCCACTAGTGTGACCTGCGAGCGCGGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGG 8

RESULT 4
BE418489/c
LOCUS
DEFINITION SCL034.E04R990601 ITEC SCL Wheat Leaf Library Triticum aestivum
CDNA clone SCL034.E04, mRNA sequence.
ACCESSION BE418489
VERSION BE418489.1 GI:9416335
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 498)
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrrells, M., Warburton, M. and Wenzel, G.
TITLE International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae

```

JOURNAL
COMMENT

Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@agr.ca
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES

source

1..498
/organism="Triticum aestivum"
/cultivar="Thatcher Lrl"
/db_xref="taxon:4565"
/clone="SCL034.E04"
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."
92 a 178 c 152 g 68 t 8 others

BASE COUNT
ORIGIN

Query Match 2.8%; Score 60; DB 10; Length 498;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgtcgacgtcgagcgccgagctccagctttgttcccttagtgagg 2141
|||||
DB 84 GATCCACTAGTGTGCGACCTGCGAGCGCGGAGCTTGTTCCTTTAGTAGGG 25

RESULT 5

BE418408/c

LOCUS

DEFINITION BE418408 567 bp mRNA linear EST 24-JUL-2000
CDNA clone SCL032.F08, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

REFERENCE

AUTHORS

"S., Dubcovsky J., Feuillet C., Gale M., Graner A., Gustafson P.,
Herrmann R.G., Holton T., Jacquemin J.M., Jia J., Joudrier P.,
Langridge P., Lazo G.R., Lin J.J., McGuire P., Ogihara Y.,
Pecchioni N., Quiset C., Schuch W., Selvaraj G., Shariflou M.,
Sorrells M., Warburton M. and Wenzel G.:
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)

JOURNAL

COMMENT

Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@agr.ca
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES

source

1..567
/organism="Triticum aestivum"
/cultivar="Thatcher Lrl"
/db_xref="taxon:4565"
/clone="SCL032.F08"
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT
ORIGIN

Query Match 2.8%; Score 60; DB 10; Length 567;
Best Local Similarity 82.1%; Pred. No. 9.5e-07;
Matches 69; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2058 tgcaggggattccctgacgatccactagtgtcgacgtcgagcgagctcc 2117
|||||
DB 107 TGCTGCTCGTCGCGAATCTTTGGATCCACTAGTGTGCGACCTGCGAGCGCGGAGCTCC 48

QY 2118 agctttgttcccttagtgagg 2141
|||||
DB 47 AGCTTTTGTTCCTTTAGTAGGG 24

RESULT 6

AI815377

LOCUS

DEFINITION AI815377 657 bp mRNA linear EST 09-JUL-1999
B0283860 Zebrafish Kidney cDNA random primed, RZPD library no: 576
Danio rerio cDNA clone RZPD clone CHBOP576F21214Q3, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AI815377 657 bp mRNA linear EST 09-JUL-1999
B0283860 Zebrafish Kidney cDNA random primed, RZPD library no: 576
Danio rerio cDNA clone RZPD clone CHBOP576F21214Q3, mRNA sequence.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Look, A.T. and Holloway, M.
Zebrafish Kidney cDNA
Unpublished (1999)
Contact: Thomas Look
Dana-Farber Cancer Institute
Pediatric Oncology Department, Mayer Building-630, 44 Binney Street
Boston, MA 02115, USA
Tel: (617) 632-5826
Fax: (617) 632-6989

The clone was obtained from an RZPD array made from an adult kidney
cDNA random primed library from Leonard Zon. RZPD library number
576. This clone is available from the Resource Centre/Primary
Database of the German Human Genome Project: RZPD (Ressourcen
Zentrum Primär Datenbank). http://www.rzpd.de
Seq primer: T7.

FEATURES

source

Location/Qualifiers
1..657
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="RZPD clone CHBOP576F21214Q3"
/clone_lib="Zebrafish Kidney cDNA random primed, RZPD
library no: 576"
/dev_stage="adult"
/note="Organ: kidney; Vector: pBK-CMV"
163 a 165 c 162 g 165 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgtcgacgtcgagcgagctccagctttgttcccttagtgagg 2141
|||||
DB 100 GATCCACTAGTGTGCGACCTGCGAGCGCGGAGCTTGTTCCTTTAGTAGGG 159

RESULT 7

BE661741

LOCUS

DEFINITION BE661741 703 bp mRNA linear EST 06-SEP-2000
138t7 GmaxSC Glycine max cDNA, mRNA sequence.

ACCESSION BE661741
 VERSION BE661741.1 GI:9987633
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max

REFERENCE 1 (bases 1 to 703)
 AUTHORS Harris,N., Chapman,B.P. and Gijzen,M.
 TITLE Gene expression in developing soybean seed coats
 JOURNAL Unpublished (2000)
 COMMENT Contact: Gijzen M
 Agriculture and Agri-Food Canada
 1391 Sandford Street, London, Ontario, Canada N5V 4T3
 Tel: 519 457 1470
 Fax: 519 457 3997
 Email: gijzenm@agr.ca.

FEATURES
 source
 1. .703
 /organism="Glycine max"
 /cultivar="Harosoy 63"
 /db_xref="taxon:3847"
 /clone_lib="Gmax5C"
 /tissue_type="Seed coats"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBK-CMV; Site.1: EcoRI; Site.2: XhoI; This cDNA library was constructed from polyA+ enriched mRNA from green seed coats in mid to late developmental stage , average fresh weight 250 mg per seed. Traces of pod and embryo tissue also present. Complementary DNA was synthesized from mRNA using an XhoI-poly(dT) linker-primer. EcoRI adapters were ligated to the blunt-ended cDNA fragments and the products were digested with XhoI for directional cloning into lambda ZAP Express vector. This lambda library was amplified once using E. coli host strain XLI Blue MRF+. Inserts were then subcloned by mass excision using ExAssist helper phage for conversion into phagemid vector pBK-CMV in E. coli host strain XL0LR."

BASE COUNT 199 a 147 c 148 g 203 t 6 others
 ORIGIN

Query Match 2.8%; Score 60; DB 10; Length 703;
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgtcgacctgcaggcgagctccagctttttgtccctttagtggg,2141
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 266 GATCCACTAGTGTGACCTGCAGCGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGGG 325

RESULT 8
 BE418461/C
 LOCUS BE418461
 DEFINITION SCL033.H02R990512 ITEC SCL Wheat Leaf Library Triticum aestivum
 CDNA clone SCL033.H02, mRNA sequence.
 ACCESSION BE418461
 VERSION BE418461.1 GI:9416307
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
 1 (bases 1 to 832)
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogiwara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.

REFERENCE 1 (bases 1 to 832)
 AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogiwara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.

International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 Contact: Cloutier S
 Cereal Research Centre, Agriculture & Agri-Foods Canada
 Winnipeg MT CANADA
 Tel: 204 983 2340
 Fax: 204 983 4604
 Email: scloutier@agr.ca
 International Triticaceae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1. .832
 /organism="Triticum aestivum"
 /cultivar="Thatcher Lrl"
 /db_xref="taxon:4565"
 /clone_lib="SCL033.H02"
 /tissue_type="Young leaf"
 /dev_stage="after 24 hour challenge with LR-AVR1"
 /note="Vector: Lambda ZAP; 1.0 kbp average insert size."

Sorrells,M., Warburton,M. and Wenzel,G.
 International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 Contact: Cloutier S
 Cereal Research Centre, Agriculture & Agri-Foods Canada
 Winnipeg MT CANADA
 Tel: 204 983 2340
 Fax: 204 983 4604
 Email: scloutier@agr.ca
 International Triticaceae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1. .832
 /organism="Triticum aestivum"
 /cultivar="Thatcher Lrl"
 /db_xref="taxon:4565"
 /clone_lib="SCL033.H02"
 /tissue_type="Young leaf"
 /dev_stage="after 24 hour challenge with LR-AVR1"
 /note="Vector: Lambda ZAP; 1.0 kbp average insert size."

BASE COUNT 179 a 249 c 226 g 141 t 37 others
 ORIGIN

Query Match 2.8%; Score 60; DB 10; Length 832;
 Best Local Similarity 100.0%; Pred. No. 1.2e-06;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgtcgacctgcaggcgagctccagctttttgtccctttagtggg,2141
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 82 GATCCACTAGTGTGACCTGCAGCGCGAGCTCCAGCTTTTGTCCCTTTAGTGAGGG 23

RESULT 9
 BE418314/C
 LOCUS BE418314
 DEFINITION SCL024.E11R990729 ITEC SCL Wheat Leaf Library Triticum aestivum
 CDNA clone SCL024.E11, mRNA sequence.
 ACCESSION BE418314
 VERSION BE418314.1 GI:9416160
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
 1 (bases 1 to 902)
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogiwara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.

REFERENCE 1 (bases 1 to 902)
 AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogiwara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.

International Triticaceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
 Unpublished (2000)
 Contact: Cloutier S
 Cereal Research Centre, Agriculture & Agri-Foods Canada
 Winnipeg MT CANADA
 Tel: 204 983 2340
 Fax: 204 983 4604
 Email: scloutier@agr.ca
 International Triticaceae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 1. .902
 /organism="Triticum aestivum"
 /cultivar="Thatcher Lrl"
 /db_xref="taxon:4565"
 /clone_lib="SCL024.E11"
 /clone_lib="ITEC SCL Wheat Leaf Library"

```

/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT      199 a      217 c      221 g      175 t      90 others
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 902;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgcacgtcgagcgcgagctccagctttgttcccttagtgagg 2141
|||||
Db 82 GATCCACTAGTGTGACCTGCGAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 23

RESULT 10
BE417925/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 934)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorreli, M., Warburton, M. and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. 934
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL012.G08"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT      211 a      271 c      222 g      224 t      6 others
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 934;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgcacgtcgagcgcgagctccagctttgttcccttagtgagg 2141
|||||
Db 82 GATCCACTAGTGTGACCTGCGAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 23

RESULT 11
BE418436/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 934)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorreli, M., Warburton, M. and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. 934
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL012.G08"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT      211 a      271 c      222 g      224 t      6 others
ORIGIN

```

```

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 960)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorreli, M., Warburton, M. and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. 960
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL033.C05"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT      230 a      234 c      236 g      214 t      46 others
ORIGIN

Query Match      2.8%; Score 60; DB 10; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2082 gatccactagtgcacgtcgagcgcgagctccagctttgttcccttagtgagg 2141
|||||
Db 82 GATCCACTAGTGTGACCTGCGAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 23

RESULT 12
BE418406/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 999)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier
, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorreli, M., Warburton, M. and Wenzel, G.
International Triticeae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. 999
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL032.F06"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT      211 a      271 c      222 g      224 t      6 others
ORIGIN

```

JOURNAL
COMMENT

Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@agr.ca
International Triticale EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES

source

1. .999
/organism="Triticum aestivum"

/cultivar="Thatcher Lr1"

/db_xref="taxon:4565"

/clone="SCL032.F06"

/clone_lib="ITEC SCL Wheat Leaf Library"

/tissue_type="young leaf"

/dev_stage="after 24 hour challenge with LR-AVR1"

/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT 281 a 180 c 256 g 244 t 38 others

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 60; DB 10; Length 999;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

2082 gatccactagtgtcgacctgcagcgccgagctccagctttgttcccttttagtgaggg 2141

||||| 81 GATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 22

RESULT 13

BE418240/c

LOCUS

BE418240 1027 bp mRNA linear EST 24-JUL-2000

DEFINITION CDNA clone SCL023.B09, mRNA sequence.

ACCESSION BE418240

VERSION BE418240.1 GI:9416086

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticaceae; Triticum.

REFERENCE 1 (bases 1 to 1027)

AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier

,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,

Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,

Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,

Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,

Sorrells,M., Warburton,M. and Wenzel,G.

International Triticale EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticaceae

Unpublished (2000)

Contact: Cloutier S

Cereal Research Centre, Agriculture & Agri-Foods Canada

Winnipeg MT CANADA

Tel: 204 983 2340

Fax: 204 983 4604

Email: scloutier@agr.ca

International Triticaceae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

Location/Qualifiers

1. .1027

/organism="Triticum aestivum"

/cultivar="Thatcher Lr1"

/db_xref="taxon:4565"

/clone="SCL023.B09"

/clone_lib="ITEC SCL Wheat Leaf Library"

/tissue_type="young leaf"

/dev_stage="after 24 hour challenge with LR-AVR1"

/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT 281 a 180 c 256 g 244 t 38 others

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 60; DB 10; Length 999;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

2082 gatccactagtgtcgacctgcagcgccgagctccagctttgttcccttttagtgaggg 2141

||||| 81 GATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 22

RESULT 13

BE418240/c

LOCUS

BE418240 1027 bp mRNA linear EST 24-JUL-2000

DEFINITION CDNA clone SCL023.B09, mRNA sequence.

ACCESSION BE418240

VERSION BE418240.1 GI:9416086

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticaceae; Triticum.

REFERENCE 1 (bases 1 to 1027)

AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier

,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,

Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,

Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,

Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,

Sorrells,M., Warburton,M. and Wenzel,G.

International Triticale EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticaceae

Unpublished (2000)

Contact: Cloutier S

Cereal Research Centre, Agriculture & Agri-Foods Canada

Winnipeg MT CANADA

Tel: 204 983 2340

Fax: 204 983 4604

Email: scloutier@agr.ca

International Triticaceae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

Location/Qualifiers

1. .1027

/organism="Triticum aestivum"

/cultivar="Thatcher Lr1"

/db_xref="taxon:4565"

/clone="SCL023.B09"

/clone_lib="ITEC SCL Wheat Leaf Library"

/tissue_type="young leaf"

/dev_stage="after 24 hour challenge with LR-AVR1"

/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT 281 a 180 c 256 g 244 t 38 others

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 60; DB 10; Length 1043;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

2082 gatccactagtgtcgacctgcagcgccgagctccagctttgttcccttttagtgaggg 2141

||||| 84 GATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 25

RESULT 15

BE418493/c

LOCUS

BE418493 1058 bp mRNA linear EST 24-JUL-2000

DEFINITION CDNA clone SCL034.E11, mRNA sequence.

ACCESSION BE418493

BASE COUNT 246 a 312 c 284 g 159 t 42 others

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 60; DB 10; Length 1043;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

2082 gatccactagtgtcgacctgcagcgccgagctccagctttgttcccttttagtgaggg 2141

||||| 84 GATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 25

RESULT 15

BE418493/c

LOCUS

BE418493 1058 bp mRNA linear EST 24-JUL-2000

DEFINITION CDNA clone SCL034.E11, mRNA sequence.

ACCESSION BE418493

BASE COUNT 246 a 312 c 284 g 159 t 42 others

ORIGIN

Query Match

Best Local Similarity 2.8%; Score 60; DB 10; Length 1043;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

2082 gatccactagtgtcgacctgcagcgccgagctccagctttgttcccttttagtgaggg 2141

||||| 84 GATCCACTAGTGTGCGACCTGCAGCGCGGAGCTCCAGCTTTGTTCCTTTAGTGAGGG 25

RESULT 15

BE418493/c

LOCUS

BE418493 1058 bp mRNA linear EST 24-JUL-2000

DEFINITION CDNA clone SCL034.E11, mRNA sequence.

ACCESSION BE418493

BASE COUNT 246 a 312 c 284 g 159 t 42 others

ORIGIN

VERSION BE418493.1 GI:9416339
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 1058)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pecchioni,N., Qaluset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticace EST Cooperative (ITEC): Production of
Expressed Sequence tags for Species of the Triticeae
JOURNAL Unpublished (2000)
COMMENT Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
International Triticace EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
FEATURES
source
1..1058
/organism="Triticum aestivum"
/cultivar="Thatcher Lrl"
/db_xref="taxon:4565"
/clone="SCL034.E11"
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT 249 a 270 c 265 g 208 t 66 others
ORIGIN

Query Match 2.8%; Score 60; DB 10; Length 1058;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2082 gatccactagtgtcgcagcgagcgagcttcttcttcccttagtgagg 2141
|||||
Db 82 GATCCACTAGTGTGCGACCTGCAGCGCGGCTCCAGCTTTGTTCCTTTAGTAGGG 23
|||||

Search completed: July 12, 2002, 13:39:55
Job time: 14437 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:47:49 ; Search time 85.28 seconds
(without alignments)
905.210 Million cell updates/sec

Title: US-09-784-554b-6

Perfect score: 3509

Sequence: 1 MKAKNSSNLSKRSKWLPPV.....SGTGYTFNHNAGDSLTMTI 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3435	97.9	695	22 AAG63964	Amino acid sequenc
2	3161	90.1	1350	22 AAG63963	Amino acid sequenc
3	3107	88.5	1352	22 AAG63962	Amino acid sequenc
4	1629.5	46.4	700	12 AAR13227	Novel endoglucanas
5	1611	45.9	617	12 AAR15241	B.lautus endol cor
6	1333	38.0	1751	20 AAR13493	Truncated cellulasa
7	236	6.7	537	18 AAW34562	Bankia gouldi endo
8	236	6.7	537	19 AAW49866	Bankia gouldi endo
9	232.5	6.6	1475	11 AAR08221	Recombinant alpha
10	211	6.0	551	18 AAW18790	Corrected Bacillus
11	186	5.3	531	16 AAW01503	60 kD endoglucanas

12	181	5.2	532	12 AAR13229	Endoglucanase enco
13	161	4.6	1684	12 AAR14948	Bacterial amylase
14	148	4.2	851	21 AAY92718	V. furnissii extra
15	135	3.8	7107	22 ABB58144	Drosophila melanog
16	134.5	3.8	5701	22 ABB36684	Peptide #4190 enco
17	134.5	3.8	5701	22 ABB22021	Protein #4020 enco
18	134.5	3.8	5701	22 AAM57448	Human brain expres
19	134.5	3.8	5701	22 AAM69843	Human bone marrow
20	134.5	3.8	5701	22 AAM05328	Peptide #4010 enco
21	134.5	3.8	26926	22 AAM05396	Human titin (conne
22	133.5	3.8	964	18 AAW32619	Cyclic-isomaltolig
23	128.5	3.7	628	13 AAR27575	ABF-A from A. nige
24	128	3.6	488	21 AAY54128	Amino acid sequenc
25	123.5	3.5	866	17 AAW02159	Soluble chitinase
26	123.5	3.5	866	21 AAY52307	Vibrio furnissii e
27	123.5	3.5	1338	14 AAR41731	High molecular wei
28	123.5	3.5	1598	18 AAW30291	Non-typeable Haemo
29	121.5	3.5	934	17 AAR92289	Cycloisomaltoligo
30	121.5	3.5	972	17 AAR92288	Cycloisomaltoligo
31	121.5	3.5	972	20 AAY32761	Bacillus CITase pr
32	121	3.4	1521	22 ABB52754	Escherichia coli p
33	121	3.4	2732	22 ABB52855	Escherichia coli p
34	120.5	3.4	699	20 AAY08471	F. balustinum CP70
35	120.5	3.4	1380	20 AAY08402	Drosophila sp. ROB
36	120.5	3.4	1381	20 AAY13564	Drosophila Robo 2
37	119.5	3.4	1282	22 ABB69127	Drosophila melanog
38	119.5	3.4	1300	22 AUA33407	Enterococcus faeca
39	119.5	3.4	1300	22 AUA35313	Enterococcus faeca
40	119	3.4	684	21 AAY59318	L. mycophilus chit
41	118	3.4	1997	21 AAB19774	Human protein tyro
42	118	3.4	1997	22 AAM78821	Human protein SEQ
43	118	3.4	2002	22 AAM78805	Human protein SEQ
44	117	3.3	2474	22 ABB22279	Novel human diagno
45	116.5	3.3	1222	21 AAB01830	H. influenzae stra

ALIGNMENTS

RESULT 1

AAG63964

ID AAG63964 standard; Protein; 695 AA.

XX

XX AAG63964;

XX

XX 29-OCT-2001 (first entry)

DT

DE Amino acid sequence of xyloglucanase enzyme.

DE

XX

XX xyloglucanase; family 44; glycosyl hydrolase; detergent;

KW

KW cellulosic fiber; textile scouring.

XX

OS Paenibacillus pabuli.

OS

XX

XX Key Location/Qualifiers

FT

FT Misc-difference 239..274

FT

FT Misc-difference 279

FT

FT

XX

XX WO200162903-A1.

XX

XX 30-AUG-2001.

XX

XX 21-FEB-2001; 2001WO-DK00116.

XX

XX 24-FEB-2000; 2000DK-0000291.

XX

XX (NOVO) NOVOZYMES AS.

XX

XX Schnorr K, Jorgensen PL, Schuelein M;

XX

XX WPI; 2001-522819/57.

DR

DR N-PSDB; AAH75061.

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries

XX Disclosure; Page 90-92; 97pp; English.

XX The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for ratting of fibers e.g. hemp, jute, flax, and
 CC linen.

XX Sequence 695 AA;

Query Match 97.9%; Score 3435; DB 22; Length 695;
 Best Local Similarity 100.0%; Pred. No. 8.2e-253;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKAKNSSNLLSKRSKWLPPVVMACIIYGGALPAPTIVVHGOTAKTVTKVDSKDRKPISP 60
 Db 1 mkaknssnllskrskwlppvmactiivggalpaptvvhgqactvtikvdsdkdrkpisp 60

Qy 61 YIYGTNQLADENLAARRLGGNRMRTGYNNWNNMNSAGSDWQSSDNFLCNGGLTKAEC 120
 Db 61 yiygtngqladgdenlaarrlgrnrmrtgynwnnnmnsnagsdwqssdnflcnggltkaec 120

Qy 121 EKPGAVTTSFHDQSLKLGAYSLVTLPMAGYVAKDNGSVQSEQAPSARQNVVNAKNAP 180
 Db 121 ekpgavttsfhdqslklgayslvtlpmagvyakdngsvqseqapsarwqvnvnaaknap 180

Qy 181 FQLQPLDNLNQVYADEFNFLVAKYGAATKAGVKGYALDNEPALWSHTHPRIHGEKVXX 240
 Db 181 fqlqpldnlngvyadefnflvlykgyaastkagvkgyaldnepalwshthprihgekvxx 240

Qy 241 XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXX 300
 Db 241 xxxxxxxxxxxx xxxxxxxxxxxx xxxxxxxxxxxx xxxxxxxxxxxx xxxxxxxxxxxx 300

Qy 301 LNSQAEKRLLDVFDVHWYPEAMGGGIRITNEVGNDETCKKARQAPRTLWDPTPKEDSWI 360
 Db 301 lnsqaekrllldvfdvhwypeamgggrritnevgnndetckkarmqaprtlwdptkyedswi 360

Qy 361 AQMNSAFLLPRLKQSDVRYYPGTLALTEYSYGGENDISGGIAMTDLVLGLGKNDVYM 420
 Db 361 aqmsafllprrlkqsdvryypgtlalteysyggendisggiantdlvlgilgkndvym 420

Qy 421 ANHWKLDGANNVSAAYKILRYNDGNATFGDISVNAQTSDIVNSSVHASVTDASYKEL 480
 Db 421 anhwkldgannvsaaykilyrnydgnatfgdisvnaqtsdivnssvhasvtdasykel 480

Qy 481 HLIVMKSMDSAFDAQFDLSGETTYSSTGKIWGFDPKNSQIKAVAPITQISGNRFTYVTP 540
 Db 481 hlivmksmdsaafaqfdlsgettysstgkiwgfdknsqikavapitqisgrfityvtp 540

Qy 541 LTAYHIVLTADNDTPPPVPEDSPESFTLRAEAGDKVDLSWDASSGVVGYSVQRATYENG 600
 Db 541 ltayhivltadndtpppvpdpesfllraeagdkvdlswdassgvvgyvqratyengp 600

Qy 601 FAAVASNLVETSDTNVNTGTSYFYKTKTAKGTSESNNVLKAVPRTPDGPDREAEAD 660
 Db 601 faavasnlvetsytdntvntgtsyfykktaktgtsesnnvlkavprtpdgpdrayaead 660

Qy 661 GTLKGTVIVSSGTFSGTGYVTNHNHAGDSLMTWTI 695
 Db 661 gtlkgktivssgtfsgtgtyvtfnfnhagdsmtmti 695

RESULT 2
 AAG63963
 ID AAG63963 standard; Protein; 1350 AA.
 AC AAG63963;
 DT 29-OCT-2001 (first entry)
 DE Amino acid sequence of xyloglucanase enzyme.
 KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
 KW cellulosic fiber; textile scouring.
 OS Paenibacillus polymyxa.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 1347
 FT /note= "Glu encoded by GAACCG"
 XX
 XX WO200162903-A1.
 XX 30-AUG-2001.
 XX 21-FEB-2001; 2001WO-DK001116.
 XX 24-FEB-2000; 2000DK-0000291.
 XX (NOVO) NOVOZYMES AS.
 XX Schnorr K, Jorgensen PL, Schuelein M;
 DR WPI; 2001-522819/57.
 DR N-PSDB; AAH75060.
 PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries
 XX Disclosure; Page 86-89; 97pp; English.

XX The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for ratting of fibers e.g. hemp, jute, flax, and
 CC linen.

XX Sequence 1350 AA;

Query Match 90.1%; Score 3161; DB 22; Length 1350;
 Best Local Similarity 86.1%; Pred. No. 1.5e-231;
 Matches 608; Conservative 18; Mismatches 56; Indels 24; Gaps 4;

Qy 1 MKAKNSSNLLSKRSKWLPPVVMACIIYGGALPAPTIVVHGOTAKTVTKVDSKDRKPISP 60
 Db 1 mkaknssnllskrskwlppvmactiivggalpaptvvhgqactvtikvdsdkdrkpisp 60

Qy 61 YIYGTNQLADENLAARRLGGNRMRTGYNNWNNMNSAGSDWQSSDNFLCNGGLTKAEC 120
 Db 61 yiygtngqladgdenlaarrlgrnrmrtgynwnnnmnsnagsdwqssdnflcngltkaec 120

QY 121 EKPQAVTTFHQDLSKLKAGYSLVTLPMAGYVAKDNGSVQESQAPSAWNOVYNKAP 180
 DB 121 ekpgavatsfhqdsikqgtyslvtlpmagvyakdngsvqesekapsarwnevnknknap 180
 QY 181 FOLOPDLNDQVYADEFNFLVKYGAASTKAGVKGYALDNEPALWSTHPRHGEKYYX 240
 DB 181 fqlpdlldngvyadefnvflvkgyvastktgkgyalndepalwsthprihgekvga 240
 QY 241 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 289
 DB 241 keldrsvslskaavdagaeifgvp-----lygfaykdltapdwnsvkngys 291
 QY 290 WFVYDYLQMLNSQAEGKRLLDVFDVHWYPEAMGGGIRITNEVGNDETCKKARMOAPRTL 349
 DB 292 wfvdylldqmlrslssaegkrlldvfdvhwypeamgggirienevndetckkarmqaprtl 351
 QY 350 WDPYTKEDSWIAQWNSAFPLPRLKQSVDKYYPGTKLALTEYSYGGENDISGGIAMTDV 409
 DB 352 wdptykedswiaqwnsefllprrlkqsvdkyypgtklalteysyggendisggiamadv 411
 QY 410 LGILGKNDVYMANWKLDGANNYSAAKLYRNYDGNKATFGDISVNAQTSDIVNSSVH 469
 DB 412 lgilgkndvymanywkldgannysaayklyrnydgnkstfgdisvhaqtsdivnssvh 471
 QY 470 ASVTDAKYKELHLIVMKNKSMDSAFDAQFDLSGETTYSYSGKTIWGDKNSSQIKAVAPITQI 529
 DB 472 asvtdasykelhlivmknksmdsafdaqfdlsgettysgktywgdknssqikeapitqi 531
 QY 530 SGNRTYTPVPLTAYHIVLTADNTPVPVPEDESFTLRABAGDGKVDLSWDASSGVVGY 589
 DB 532 sgnrttytvppltayhivltadntpvpvpepdesftlrabagdgk-vsiyldassgvvy 587
 QY 590 SVQRATYENGPFPAVASNLVETSYDTNVTNGTSYYKYITAKTACGSESNLKAVPRTP 649
 DB 598 svqratyengpfpaavasnlvetsydtntvngtsyykyitaktgtcsaenlkavprap 647
 QY 650 VDGPRDYEADGTLLKGTIVESSGTFSGTGYVTNPFHAGDSLTMTI 695
 DB 648 vdgpdyeeadgtllkgtvessgtfsgtgyvtnpfhagdsitmti 693

RESULT 3

AAG63962
 ID AAG63962 standard; Protein; 1352 AA.

AC AAG63962;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of xyloglucanase enzyme.

KW xyloglucanase; family 44; glycosyl hydrolase; detergent;

KW cellulosic fiber; textile scouring.

OS Paenibacillus polymyxa.

FN W0200162903-A1.

PD 30-AUG-2001.

PF 21-FEB-2001; 2001WO-DK00116.

PR 24-FEB-2000; 2000DK-0000291.

PA (NOVO) NOVOZYMES AS.

PI Schnorr K, Jorgensen PL, Schuelein M;

DR WPI; 2001-522819/57.

DR N-PSDB; AAH75059.

PT New xyloglucanase enzyme belonging to glycosyl hydrolases family.

PT useful for detergent compositions, and textile or cellulose fiber
 processing industries
 XX
 PS Example 3; Page 81-85; 97pp; English.
 XX

CC The present sequence represents a xyloglucanase of the invention. The
 specification describes a xyloglucanase enzyme belonging to family 44
 of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 of at least 30% at pH 5-8. The enzyme exhibits high performance in
 detergent compositions and prevents binding of certain soils to the
 xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 room temperature and has a half life of more than 50 days when incubated
 in a full formulated liquid detergent at 30 plus degrees celsius. The
 enzyme is used in detergent compositions, textile industry for improving
 the properties of cellulosic fibers, yarn, and woven or non-woven
 fabrics, preferably in textile scouring process, and in cellulose fiber
 processing industry for retting of fibers e.g. hemp, jute, flax, and
 linen.

XX Sequence 1352 AA;

Query Match 88.5%; Score 3107; DB 22; Length 1352;
 Best Local Similarity 83.9%; Pred. No. 2e-227;
 Matches 592; Conservative 28; Mismatches 64; Indels 22; Gaps 3;

QY 1 MKAKNSNLSKRSKWLPPVVMACTIIVGGALPAPTIVHGGTAKTVTKVDTSKDRKPISP 60
 DB 1 mraknsnllfkfkskwpvmactmivggalpavvhggtaktitkvdtkdrkpisp 60

QY 61 YIYGTQDLADGDNLAARLGGNRWFGYWNWNSNAGSDQSSDNFLCNGGLTKAC 120
 DB 61 yiygtqdladgdnlaarlggnrwtgynwnnsmnagsdwqssdnylcsnggltkaac 120

QY 121 EKPQAVTTFHQDLSKLKAGYSLVTLPMAGYVAKDNGSVQESQAPSAWNOVYNKAP 180
 DB 121 ekpgavtsfhqdsikqgtyslvtlpmagvyakdngsvqesekapsarwnevnknknap 180

QY 181 FOLOPDLNDQVYADEFNFLVKYGAASTKAGVKGYALDNEPALWSTHPRHGEKYYX 240
 DB 181 fqlpdlndnrvyadefnvflvkgyastkagvgyaldnepalwsthprihgekvga 240

QY 241 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 289
 DB 241 keldrsvslskaavdagaeifgvp-----lygfaykdltapdwnsvkngys 291

QY 290 WFVYDYLQMLNSQAEGKRLLDVFDVHWYPEAMGGGIRITNEVGNDETCKKARMOAPRTL 349
 DB 292 wfvdylldqmlrslssaegkrlldvfdvhwypeamgggirienevndetckkarmqaprtl 351

QY 350 WDPYTKEDSWIAQWNSAFPLPRLKQSVDKYYPGTKLALTEYSYGGENDISGGIAMTDV 409
 DB 352 wdptykedswiaqwnsefllprrlkqsvdkyypgtklalteysyggendisggiamtdv 411

QY 410 LGILGKNDVYMANWKLDGANNYSAAKLYRNYDGNKATFGDISVNAQTSDIVNSSVH 469
 DB 412 lgilgkndvymanywkldgannysaayklyrnydgnkstfgdisvhaqtsdivnssvh 471

QY 470 ASVTDAKYKELHLIVMKNKSMDSAFDAQFDLSGETTYSYSGKTIWGDKNSSQIKAVAPITQI 529
 DB 472 asvtdasykelhlivmknksmdsafdaqfdlsgettysgktywgdknssqikeapitqi 531

QY 530 SGNRTYTPVPLTAYHIVLTADNTPVPVPEDESFTLRABAGDGKVDLSWDASSGVVGY 589
 DB 532 sgnrttytvppltayhivltadntpvpvpepdesftlrabagdgkvdlswdassgvvy 589

QY 590 SVQRATYENGPFPAVASNLVETSYDTNVTNGTSYYKYITAKTACGSESNLKAVPRTP 649
 DB 590 svqratyengpfpaavasnlvetsydtntvngtsyykyitaktngkssesnlkavprtp 649

QY 650 VDGPRDYEADGTLLKGTIVESSGTFSGTGYVTNPFHAGDSLTMTI 695
 DB 650 vngprdyeeadgtllkgtivessgtfsgtgyvtnpfhagdsitmti 695


```

:|: |::| |::| |::| |::| |::| |::|
Db 1394 rssaatgikgylidnepdlwftthprhpqkvtcscelinksvlakviktlpdaefgp 1453
Qy 267 XXXXXXXXXKLHXT--GTSVKGNYFVYLDQMLNSQAEGKRLLDVDFVHWYPEAMG 324
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1454 asyfvgylltldapdwngvkgnhwflswylegmkkasdsfgkrlldvldihwybeagv 1513
Qy 325 GGIRITNEVGNDETHK---ARMQAPRTLDWDPYK-----EDSWTAOWNSAFLPL 370
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1514 ggvricfd-genstsrdaalrmqaprtlwdpykttqkggitagenuswlnqwfpeylpl 1572
Qy 371 LPRLKQSDVXYPGTKLALTEYSYGGENDISGGTAMTDVLGKNDVYMANWYKLDGA 430
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1573 lpnkadidkypptkilaitefygdkhsglaladvlgfgygvyymaarw---gds 1629
Qy 431 NNYVSAAYKLYRNDGKNATFGDISVNAQSDIVNSSVHASVTDASKEHLHLYMKNMSMD 490
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1630 gsyaaaynlynydgkgsrystcvsaettdvenmpvyasiegeddstvhiilinnrd 1689
Qy 491 SAFAQDFDLGSETTYSKGIWGFDPKNSQIKAVAPITQISGNRTYTPPLTAYHIVLTA 550
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1690 rklkaelkmntrvtytgelygffdstssqirkmgvisnignnttitievpltyhivlts 1749

```

RESULT 7

```

AAW34562
ID AAW34562 standard; Protein; 537 AA.
XX AC AAW34562;
XX DT 12-MAR-1998 (first entry)
XX DE Bankia gouldi endoglucanase.
XX KW Glycosidase; thermostable; textile; food processing; pharmaceutical;
XX KW detergent; baking; industry; Thermococcus; Staphylothermus;
XX KW Pyrococcus; glucose; soluble oligosaccharide; endoglucanase.
XX OS Bankia gouldi.
XX PN WO9725417-Al.
XX PD 17-JUL-1997.
XX PF 10-JAN-1997; 97WO-US00092.
XX PR 13-SEP-1996; 96US-0712612.
XX PR 11-JAN-1996; 96US-0583787.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

```

```

PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
DR WPI; 1997-372858/34.
DR N-P5DB; AAT93686.
XX
XX New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
PT and Pyrococcus, used in the textile, food processing,
PT pharmaceutical, detergent and baking industries
XX
XX Claim 4; Fig 9; 82pp; English.
XX
XX The present sequence represents endoglucanase isolated from Bankia
CC gouldi. The enzyme or its encoding nucleic acid sequence is used
CC for generating glucose from soluble oligosaccharides. The enzyme can be
CC used in the food processing, pharmaceutical, textile, detergent and
CC baking industries. The enzyme is also used to treat lactose intolerance,
CC as a diagnostic reporter molecule, in corn wet milling or in the fruit
CC juice industry. The enzymes can be used to hydrolyse guar gum to remove
CC non-reducing terminal mannose residues. The nucleic acids encoding the
CC enzyme may be used to generate probes to identify similar sequences.
XX
XX Sequence 537 AA;

```

Query Match

6.7%; Score 236; DB 18; Length 537;

Best Local Similarity 21.2%; Pred. No. 1.8e-09;

Matches 131; Conservative 77; Mismatches 237; Indels 174; Gaps 24;

```

Qy 20 VMACTIIVGGALPAPTIVHGTAKTVTKVDTSKDRKPISPYIYCTNODLAGDENLA--- 76
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 3 irlatlaicalcaalspvtf-----adnvtvgldadggkklisralygmnnnsna--esltdtd 55
Qy 77 -----ARRLGGNRMTGYNWENNNSNAGSDMQQ---SSDNFLCNNGGLTKAECEK 122
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 56 wqfrdagvrmilrenggnnstkwnqlhlis-shpdvnnvyagnnannrhrvaligenl-- 112
Qy 123 PCAVATTSFHDQSLKLGAYSLVTLPMAGVYAKDNGSGVSEQAPARWNVVNAKNAPFQ 182
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 113 pgadt-----mwaqgli-----gkvaatsaynfndwefnqswtgv--aqnlagg 156
Qy 183 LQPDLN-DNQVYADEFVNFVKKYGAASATKA-----GVKK-----YALDNEPAL 225
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 157 gepnldgggealvegdpnljlmwspadtvgllhdfgvgnglgvrrgkakywsmndnepgi 216
Qy 226 WSHTHPRIHGEKVVXXXXXXXKXXXXXXXXXXXXXXXXXXXXXKFLHXIGTSVK 285
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 217 wvghddvvekq-----tpvedflhtyfetakkarakfpgikitgvpv 259
Qy 286 GN-----YSW-----FVDYVLDQMLNSQAEGKRLLDVDFVHWYPEAMGGGIR 328
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 260 anewqwyawggfsvpqqeqgfmsmemyfkrvseeqragvrlldvldlhypgaya--- 316
Qy 329 ITNEVGNDETTRKARMOAPRTLDMPY-----KEDSWTAQWNSAFLPLPRLKQSV 378
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 317 -----edivqlhrtffdrfdvsiidangvkmveggwddsinkey--ifgrvndwl 363
Qy 379 DKXY---PCTKLALTEYSYGGENDISGGTAMTDVLGKNDVYMANWYKLDG----- 429
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 364 eeymgpdhgtigtitemcvrnnpmtaltaiwysamlgtfadngveiftpcwntgmwetlh 423
Qy 430 -----ANNVSAAYKLYRNDGKNATFGDISVNAQSDIVNSSVHAS--- 471
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 424 lfsrynkpvrassssaleefvsa---ysaineadamtvllvnrstsethtatvalddf 479
Qy 472 VTDASVKEHLHLYMKNMSDMSAFDAQDFDLGSETTYSKGIWGFDPKNSQIKAVAPITQISG 531
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 480 pldgpvrtirl-----hnlpgveetfvshrdnaiekgt-----vrsad 516
Qy 532 NRFTYTPPLTAYHIVLTA 550
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 517 nvtlelplpsvtaililka 535

```

RESULT 8

```

AAW49866
ID AAW49866 standard; Protein; 537 AA.
XX AC AAW49866;
XX DT 21-DEC-1998 (first entry)
XX DE Bankia gouldi endoglucanase 37GP1.
XX KW Glycosidase; 37GP1; thermostable enzyme; oligosaccharide;
XX KW glucose; sugar; baking; textile; detergent; endoglucanase.
XX OS Bankia gouldi.
XX PN WO9824799-Al.
XX PD 11-JUN-1998.
XX PF 08-DEC-1997; 97WO-US22623.
XX PR 10-OCT-1997; 97US-0949026.

```


RESULT 11

AAW01503
ID AAW01503 standard; protein; 531 AA.XX
AC AAW01503;XX
DT 26-FEB-1997 (first entry)XX
DE 60 kD endoglucanase, EG C.XX
KW Detergent composition; cellulase; retaining-type activity; catalytic
KW activity; cellotriase; particulate soil removal; colour clarification;
KW cleaning; cellulose-containing fabric; cellulohydrolase; endoglucanase;XX
OS Bacillus lautus, NCIMB 40250.XX
PN W09502675-A1.XX
PD 26-JAN-1995.XX
PF 07-JUL-1994; 94WO-DK00280.XX
PR 11-OCT-1993; 93DK-0001135.XX
PR 12-JUL-1993; 93EP-0870131.XX
PA (NOVO) NOVO-NORDISK AS.XX
PA (PROC) PROCTER & GAMBLE CO.XX
PI Convents AC, Jeffreys B, Schuelein M, Tikhomirov DF;XX
DR WPI; 1995-067325/09.XX
PT Detergent compsn. contg. two cellulase components - the first
PT removing soil particles and the second capable of colour
PT clarification, useful in laundry compsns.XX
PS Claim 28; Page 68-70; 83pp; English.XX
CC Detergent compositions comprising: 1) a first cellulase component having
CC retaining-type activity, pref. having a catalytic activity on cellotriase
CC at pH 8.5 corresponding to that of at least 0.01 s<-1>, and capable of
CC particulate soil removal; and 2) a second cellulase component having
CC multiple domains comprising at least one non-catalytic domain attached to
CC a catalytic domain, pref. having a catalytic activity on Red Avicel 7.5
CC per 1 mg of cellulase protein higher than 10<-4> IU and being capable of
CC colour clarification, where at least one of the cellulase components is a
CC single (recombinant) component, are useful for cleaning and colour
CC clarification of cellulose-containing fabrics. The second cellulase
CC component can be an endoglucanase which is immunoreactive with an
CC antibody raised against a highly purified -60 kD endoglucanase
CC derived from Bacillus lautus, NCIMB 40250, and is esp. the present
CC sequence, designated EG C.XX
SQ Sequence 531 AA;

Query Match 5.3%; Score 186; DB 16; Length 531;

Best Local Similarity 38.2%; Pred. No. 1.1e-05;

Matches 39; Conservative 17; Mismatches 36; Indels 10; Gaps 2;

QY 567 LRAEAGDKVDLSMDASSGVGVSVORATYENGPFPAVASNLVETSYDTNVTNGTSYY 626

Db 345 lsangnagqvsltnavsgatsytkrattsggpytnvdrgrvatsytpgtlmgtttYYY 404

QY 627 KITAKTKAGTSNSNLKAVPRTPVDPDRYEAEDELTKGTIV 668

Db 405 vvransagssansaqasa--tpasg-----gastgnlv 436

RESULT 12

AAR13229

ID AAR13229 standard; Protein; 532 AA.

XX
AC AAR13229;XX
DT 14-OCT-1991 (first entry)XX
DE Endoglucanase encoded by endo3 gene.XX
KW Cellulase activity; detergent.XX
OS Bacillus spp. NCIMB 40250.XX
FH Key Location/Qualifiers

FT Cleavage-site 36..37

FT Peptide 1..36

FT /label= signal peptide

FT Peptide 37..531

FT /label= mature peptide

XX
PN W09110732-A.XX
PD 25-JUL-1991.XX
PF 18-JAN-1991; 91WO-DK00013.XX
PR 19-JAN-1990; 90DK-0000164.XX
PA (NOVO) NOVO NORDISK A/S.XX
PI Jorgensen PL, Schuelein M, Hansen C;XX
DR WPI; 1991-238020/32.XX
DR N-PSDB; AAQ13003.XX
PT Enzyme exhibiting cellulase activity from Bacillus sp. - is an
PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
PT fabrics.XX
PS Claim 1; Page 80; 96pp; English.XX
CC The enzyme is encoded by a 11000 bp. EcoRI fragment of Bacillus
CC spp. PI2336 DNA contained in plasmid pPL591. It exhibits an
CC endoglucanase activity of at least 10 (pref. at least 25)
CC carboxymethyl cellulose (CMC) endoase units per mg total protein
CC under alkaline conditions. It is especially useful as a
CC cellulolytic agent and has been found to be more stable during
CC washing (60 mins. at 40 deg.) in the presence of conventional
CC detergents than a commercial cellulase preparation. It may also
CC show increased storage stability in liq. detergents contg.
CC proteases. The sequence was deduced from the DNA (AAQ13003), it is a
CC product of the endo3 gene. See also AAR13227 and AAR13228.XX
SQ Sequence 532 AA;

Query Match 5.2%; Score 181; DB 12; Length 532;

Best Local Similarity 40.2%; Pred. No. 2.7e-05;

Matches 41; Conservative 16; Mismatches 35; Indels 10; Gaps 3;

QY 567 LRAEAGDKVDLSMDASSGVGVSVORATYENGPFPAVASNLVETSYDTNVTNGTSYY 626

Db 345 lsangnagqvsltnavsgatsytkrattsggpytnvdrgrvatsytpgtlmgtttYYY 404

QY 627 KITAKTKAGTSNSNLKAVPRTPVDPDRYEAEDELTKGTIV 668

Db 405 vvransagss-ansaqasar--pasg-----gastgnlv 436

RESULT 13

AAR14948

ID AAR14948 standard; Protein; 1684 AA.

XX
AC AAR14948;

XX 25-FEB-1992 (first entry)
 XX Bacterial amylase A-180.
 XX Maltopentose; G5; starch hydrolysis.
 XX EP459385-A.
 XX 04-DEC-1991.
 XX 28-MAY-1991; 91EP-0108669.
 XX 31-MAY-1990; 90DE-4017595.
 XX (CONE) CONSORT ELEKTROCHEM IND.
 XX Schmid G, Candussio A, Bock A;
 XX WPI; 1991-355676/49.
 XX N-PSDB; AAQ13939.
 XX New bacterial amylase, A-180 for malto: pentose prodn. - by
 PT hydrolysis of starch, providing high yield and modifiable for
 PT secretion from host cells
 XX Disclosure; Page 7; 21pp; German.
 XX The amino acid sequence is that of bacterial amylase A-180 which is
 CC used to prepare maltopentose (G5) in high yields by hydrolysis of
 CC starch. It can be modified to ensure its excretion, obviating the need
 CC to concentrate and purify the enzyme, i.e. the culture supernatant can
 CC be used directly for G5 prodn. G5 yields of over 90% are possible,
 CC eliminating the need for further G5 purification. G5 is used in medical
 CC assays for diagnosis, as carbohydrate sources in liquid feeding compns.
 CC and for converting (by esterification) fatty acids to water-soluble
 CC form, suitable for use in stable infusion solns.
 XX Sequence 1684 AA;
 XX
 Query Match 4.6%; Score 161; DB 12; Length 1684;
 Best Local Similarity 20.6%; Pred. No. 0.0046;
 Matches 146; Conservative 62; Mismatches 256; Indels 244; Gaps 29;
 QY 64 GTNQDLAGENLAARRLG-----GNRMGTGYNNWNNMNAQSDWQ---QSSDNFLCNN 112
 DB 461 ggvgvfygde--tarplgdggsdpeqgtr-ssmwananinqvishwqklgqfrnnhlaig 517
 QY 113 GGLTKAECEKPGAVTTFSDHQSILGLAYSLVLPAGYVAKDNGSVQSEEQAPSAWNO 172
 DB 518 agahqklsdpspytfartyesddi-----ydevvatgaggtt 554
 QY 173 VYNAKNAPFQLPDLNDNOVYADEFV--NFLVKYKGAATKAGVKGYALDNEPALWSHTH 230
 DB 555 avtegv-----fedgtvrdytagdettvk-gtatftagtgqillenta----- 600
 QY 231 PRIHGEKYYVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXITGTSVKGNYSW 290
 DB 601 -----epvtnl-----pivsatpgnssf 618
 QY 291 FVDYLDQRLNSAQEKRLLDVDFVHVPAMGG-GIRITVEGNDETKKARMQA----- 345
 DB 619 rtdtitinvdramgkytldgsdpadgltfmdgeelvigadmedetatlriyaenen 678
 QY 346 -----PRTLWDPTPKEDSWIAQWNSAFLPLPRLKQSVBKYPGT----- 385
 DB 679 girtsytrkvdpdallevyfkp---adwtgphi-----yydtfpepev 723
 QY 386 -----KLALTE---YSGGENDISGGIAMDVLG---ILGKNV-----YMANWK 425
 DB 724 twttapentlveddwyvvyfenaesaniifkdssgkqpgpnepgffidqigwydgkvkl 783

QY 426 LKDGANNVSAAYKLYRNYDGKNATFGDISVNAQTS-----DIVNSVHASV 472
 DB 784 dsdpferpekpattpknlsvvntettvtfewdgsdgyvveyeilldedvvastirttf 843
 QY 473 TDASYKELHLIVNKSMDSPADQFDLSGETTYSYSGKLTWGFKNSSQIKAVAPITQISGN 532
 DB 844 tde-----dlpddty-----tysvvavgeggqksa- 869
 QY 533 RFTYTPPLTAYHIVLTADNDTPVPVDEPSESFTLRAEAGDGKVDLSWDSSG-VVGYSV 591
 DB 870 -----psealkvtlleendepkaepenlri-aditdtvtinwnasngyvtgyev 921
 QY 592 ORATYENGPFPAVASNLVETSTDTNVTNGTSYYKTKAKTKAGTSESNVLKAVPRTPVD 651
 DB 922 lrd-----gvligettrtftidgldadrttytivalgdgg-qkadpsealevtqtge 973
 QY 652 GPD-----RYEADGTLKGTIV-----ESSGTGFS 676
 DB 974 kpegnlvtlyykgfdtpymhyrpegg--ewclvpgirmeeselagys 1019

RESULT 14

AA92718
 ID AAY92718 standard; Protein; 851 AA.

XX
 AC AAY92718;

XX
 DT 29-AUG-2000 (first entry)

XX
 DE V. furnissii extracellular chitinase.

XX
 KW E-chitinase; extracellular chitinase; hydrolysis; oligosaccharide;
 KW chitin; N,N'-diacetylchitobiose; restriction-modification system.

XX
 OS Vibrio furnissii.

XX
 PN WO200024874-A1.

XX
 PD 04-MAY-2000.

XX
 PF 27-OCT-1999; 99WO-US25180.

XX
 PR 28-OCT-1998; 98US-0106047.

XX
 PA (UYJO) UNIV JOHNS HOPKINS.

XX
 PI Fomenkov A, Keyhani NO, Roseman S;

XX
 DR WPI; 2000-350720/30.

XX
 DR N-PSDB; AAA09500.

XX
 PT Isolated polynucleotide for producing chitin oligosaccharides encodes
 PT extracellular chitinase

XX
 PS Claim 22; Fig 4; 95pp; English.

XX
 CC This is the Vibrio furnissii extracellular chitinase (E-chitinase).
 CC The E-chitinase is capable of hydrolyzing chitin in a standard chitinase
 CC assay. A Vibrio furnissii null mutant lacking E-chitinase activity as
 CC determined by a standard or cell-mediated chitinase assay is also
 CC claimed. The E-chitinase can be used to produce a pure preparation of
 CC N,N'-diacetylchitobiose ((GlcNAc)₂). Methods of transforming prokaryotic
 CC cells with or transconjugating a recombinant vector in which the cells
 CC comprise a restriction-modification system are claimed. A recombinant
 CC Escherichia coli strain comprising a sequence comprising a V. furnissii
 CC DNA methylase gene is also claimed. The polynucleotides and polypeptides
 CC are useful for producing chitin oligosaccharides for research, medical,
 CC agricultural or commercial use.

XX
 SQ Sequence 851 AA;

Query Match 4.2%; Score 148; DB 21; Length 851;

Best Local Similarity 19.9%; Pred. No. 0.017;
Matches 151; Conservative 88; Mismatches 304; Indels 214; Gaps 37;

QY 42 AKTIVTIKVDTSKDRKIPISPIYGTNODLAGDENLAARRL-----GGRMTGYWNENN 93
Db 13 asialalsgsalaapeapsi-----dvysnnlqfksielamettagymvthydeap 67

QY 94 MSNAGSDW-----QSSDNFLCNGGLTKAE-----CEKPGAV 126
Db 68 isitfqwsdetgntykiyfdgeevagsgisqstasfyskgylyleaacddsgcs 127

QY 127 TTSFHDQSLKLGAYSLVTLPMAGYVAKDNGSGVQSEQAQARWQVNAKNAPFQLOPD 186
Db 128 tsa-----pveltia-dtdg-----ahikpilmnvd 152

QY 187 LNDNOVYADEFNFLYKKGAASTRAGYKG--YALDNEPALMSHHPRIHG-----E 236
Db 153 pn-nktyttcd-antvv---gtyfvewgylgrnytdvnpa-qnlth-ilygfipicgpn 205

QY 237 KVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKL-HXIGTSVKGNYS----- 289
Db 206 svksvgnsynalmtacqgpydevihdpwaaykksfaqagheyspikgnymmmalk 265

QY 290 -----W-----FVYYLDQRLMSQAEGKRL-----DVFDVHW-YPEAM 323
Db 266 qrypdkliipsvgvgtlsqdfdfctkanrdtfvasvkrflttwkffgvdidwefpgd 325

QY 324 GGGIRITNEVGNDKTKARMAQAPRLWPTYKEDSWIAQWNSAFPLPLRLKQSDVK--- 380
Db 326 gaapdigdpvngpypaylmgqmrlndlglaetgrtyeltsa-----lgvgdykied 378

QY 381 -----YYPGTKALTEYSYGGENDISG-----GIAM-----TDVLGILGKNDVYMAN 422
Db 379 vnyadavqymdyifamtydfyggwnnvighqtaIncgfmrpgqcdgtgvdengeaygpp 438

QY 423 YWKLKDG-----ANNYSAAKLYRNYDG-KNATFGD-----ISVNAQTSDIV 464
Db 439 aytadrgtqlllagvpankvlgtamygrgdwgvmpssldpdmgtvgvngklgtsta 498

QY 465 NSSVHASVTDASYKELHLIVMKNMSDAFDAQFDLSGETTY-----SSCKINGFDKNSQI 520
Db 499 qgvweagvldykgikannlmgadntgingfeygydaqaepwvnrsgqlitfddrs-v.557

QY 521 KAVAPITQISG--NRFYTVPP-----LTAYHIVLTADNDTP-----VPPVED 561
Db 558 kakgavarslglagifswaidadngilnamhegl-adggtppanraptanagvaqtvg 616

QY 562 PESETLRAEA---GDGV-DLSWDASSGVGVSVQRATYENGPPFAAVASNLVETSYTDTN 617
Db 617 pasvtldgsasdsdgtiasylweqtgg-----davtln-----asavqasftape 663

QY 618 VINGTSYKYKITAKTKAGTSSESNVLKAVPRTPVDGPD 654
Db 664 vtseqftfkltvdddgasas-attvtvqpvga 699

RESULT 15
ABB58144
ID ABB58144 standard; Protein; 7107 AA.
AC ABB58144;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1224.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
FA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR N-PSDB; ABL02247.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 1224; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 7107 AA;

Query Match 3.8%; Score 135; DB 22; Length 7107;
Best Local Similarity 18.1%; Pred. No. 3.5; Indels 278; Gaps 33;
Matches 137; Conservative 86; Mismatches 255;

QY 34 PTVVHGO-----TAKTIVTIKVDTSKDRKIPISPIYGTNODLAGDENLAARRLGGNRTG 87
Db 4670 psikegeysfrvvaenevgrsdpskpsqit-----ieeqnkpkmelgkyr--- 4717

QY 88 YNWNENMSNAGSDWQ-----QSSDNFLCNGGLTKAECEKPGAVTYSFHD 132
Db 4718 -----divrcgddfsihvpylafpknafwysndmlddnrvhkhltddaasvvyk-hs 4772

QY 133 QSLKLGAYSLVTLPMAGYVA-----KDGNG 157
Db 4773 kradsggyrlqlkntsgfdtatinrvldrpsoptrlradefsgdsaltlywnppnddgs 4832

QY 158 SVQ-----ESEQAPSARWQVNAKNAPFQLOPLNDNQVYADEFNFLVKKYGAATKAG 213
Db 4833 alqnyliekearsstskvsfctvfvirnlvlnkey--dfvrvaenkyqsdp--- 4887

QY 214 VKGYALDNEPALMSHHTH--PRIHGEKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 271
Db 4888 -----antsepilarhpfidpntpg----- 4907

QY 272 XXXFKLHXIGTSVKGNYSWFVDDYLDQRLMSQAEGKRLLDVDFVHWYPEAMGGGIRITN 331
Db 4908 -----iphgidste-----dsitiaetkpkhdggsptg 4936

QY 332 EVGNDETKARMAQAPRLWPTYKEDSWIAQWNSAFPLPLRLKQSDVKYYPGTKL----- 387
Db 4937 yi-----iekrllsdd-----kwtkavhalcpdlsciplienaeeyfrv 4977

QY 388 -----ALTEYSYGGEND-----ISGGTAMTDVLGILGKNDVYMANV----- 423
Db 4978 aavnaagqaysgssdlifcrrpphapktsdlsrdmvtviagdefrcvpyhasprpta 5037

QY 424 -W-----KLKDGANNYSAAKLYRNYDGKNATFGDTSV---NAQTSDIVNSSV 468
XX

Db 5038 swslngrlevipgerikfidsndyas-----myynksakrdetgsytlitltnnkgdsd--tasc 5091
Qy 469 HASVTDA-----SYKELHLIVMNKSMDSAFDAQDLSGE--TTYSSGKIWGFDKNS 517
Db 5092 hvtvdrplppggplnayd-----itpdtctlawktplddggspitnyvveki---dnsg 5143
Qy 518 SQIKAVAPITQISGNRFTYTV---PPLTAYHIVLTADND-----TPVPPVEDPESFTL 567
Db 5144 swvk-----issfvrnth-ydvmgplephykynfrvraenqygsdpldilepivakqftv 5198
Qy 568 RAEAG-----DGKVDLSW-----DASSGVVGYSVQ-RATYENGPFPAVASNLVETSY 613
Db 5199 pdepqpkvidwdsgnvtllwtrplsdgggriggyqieyrdilndsswnaydylikdky 5258
Qy 614 TDTNVTNGTSYYKITAKTSGTESNVLKAVPRTP 649
Db 5259 qlynlingseyefrikaknaaglsk-----pspsp 5287

Search completed: July 13, 2002, 00:47:54
Job time: 21610 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:49:42 ; Search time 35.21 seconds
(without alignments)
482.130 Million cell updates/sec

Title: US-09-784-554B-6
Perfect score: 3509
Sequence: 1 MKAKNSSNLSKRSKWLPPV.....SGTGYVTNHNAGDSLMTI 695

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PTCUTS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1629.5	46.4	700	2	US-07-862-588B-2
2	1612	45.9	617	1	US-08-361-920-29
3	1612	45.9	617	1	US-08-479-939-29
4	1612	45.9	617	1	US-08-483-432-29
5	1333	38.0	1751	4	US-09-136-574A-44
6	211	6.0	551	2	US-09-033-537A-1
7	186	5.3	531	2	US-07-862-588B-7
8	128.5	3.7	628	2	US-07-952-853-22
9	128.5	3.7	628	2	US-08-914-848-22
10	123.5	3.5	866	1	US-08-386-727-8
11	123.5	3.5	866	2	US-08-600-452A-8
12	123.5	3.5	1338	4	US-08-728-470-9
13	123.5	3.5	1599	2	US-08-617-697-9
14	123.5	3.5	1599	2	US-08-617-697-10
15	120.5	3.4	1381	4	US-09-540-245A-16
16	114	3.2	1529	2	US-08-728-470-10
17	114	3.2	1529	4	US-08-719-641-10
18	114	3.2	1600	2	US-08-617-697-10
19	112	3.2	509	3	US-08-822-324-8
20	112	3.2	750	3	US-08-814-052-2
21	112	3.2	750	3	US-08-812-829-2
22	111.5	3.2	1358	1	US-08-404-665-4
23	111.5	3.2	1358	1	US-08-404-671-4
24	111.5	3.2	1358	1	US-08-404-781-4
25	111	3.2	942	1	US-08-141-324-14
26	111	3.2	942	1	US-08-541-902-14
27	108.5	3.1	1536	1	US-08-038-682-2

28	108.5	3.1	1536	1	US-08-302-832-2	Sequence 2, Appli
29	108.5	3.1	1536	2	US-08-530-198-2	Sequence 2, Appli
30	108.5	3.1	1536	2	US-08-469-880-2	Sequence 2, Appli
31	108.5	3.1	1536	2	US-08-728-470-2	Sequence 2, Appli
32	108.5	3.1	1536	2	US-08-617-697-2	Sequence 2, Appli
33	108.5	3.1	1536	4	US-08-719-641-2	Sequence 2, Appli
34	108.5	3.1	2324	1	US-08-283-857-1	Sequence 1, Appli
35	108.5	3.1	2324	5	PCT-US95-09819-1	Sequence 1, Appli
36	108	3.1	1912	1	US-08-409-995-4	Sequence 1, Appli
37	108	3.1	1912	3	US-08-685-467-4	Sequence 4, Appli
38	108	3.1	2353	4	US-09-377-155-33	Sequence 33, Appli
39	108	3.1	2353	4	US-08-913-942-4	Sequence 4, Appli
40	108	3.1	2353	4	US-09-669-974-33	Sequence 33, Appli
41	108	3.1	2411	4	US-09-268-347-36	Sequence 36, Appli
42	107.5	3.0	2327	6	5455158-1	Patent No. 5455158
43	107	3.0	1911	1	US-08-348-006B-5	Sequence 5, Appli
44	107	3.0	1911	2	US-08-800-825A-5	Sequence 5, Appli
45	107	3.0	1911	4	US-09-158-657-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 59167960 NO. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 46.4%; Score 1629.5; DB 2; Length 700;
Best Local Similarity 48.8%; Pred. No. 2.6e-138;

Matches 329; Conservative 84; Mismatches 198; Indels 63; Gaps 9;

```

Qy 11 SKRSKLPVVMACTIIVGGALPAPTVHGGTAKTVIKVDTSKDRKPIPIYIGTQDIA 70
Db 3 TRQKRLFLVSAALAVSLTMTVPASVNA--AASDVFTINTQSERAAISPIYIGTQDLS 61
Qy 71 GDEWLAARRLGGNMTYNNWNNMNSAGSDQSSDNFLCNGGLTKACEKPGAVTTSF 130
Db 62 GTENWSSRRLLGGNRLTGYNNWNNMNSAGSDQSSDNFLCNGGLTKACEKPGAVTTSF 121
Qy 131 HDQSILKLGAYSLVTLPMAGYVAKDNGSVQESQAQSRNNQVNNAKNAPFQLOPDLND 190
Db 122 HDKSLNGAYSVTLQMGAYVSRDKNGPVDESETPRWDKVEFAKNAPFSLQDNLNDG 181
Qy 191 QVYADEVNFVLKYGAASTKAGYKGYALDNEPALMSHTPHRIGEBKVVXXXXXXX 250
Db 182 QYVWDEEVNVLNRYGNASTSTGKAYSLDNEPALMSETHPRHPQLQAAELVAKSIDL 241
Qy 251 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXI--GTSVKGNYSWFDVYDQLQRLNSQAE 308
Db 242 SKAVKNVDPHAEIFGPAFYGLYSLQDAPDWPSLQGNYSWFDVYDQLQRLNSQAE 301
Qy 309 RLDDVDFVHHYPEAMGGGIRIT-NEVCNDETKKARQAPRTLWDPTYKEDSWIAQNSAF 367
Db 302 RLDDVDFVHHYPEAMGGGIRIT-NEVCNDETKKARQAPRTLWDPTYKEDSWIAQNSAF 361
Qy 368 LPLRLKQSDVKYYPGTKLALTEYSYGGENDISGGTAMTDLGLGKNDVYMANWKLK 427
Db 362 LPLPKLQSSIQIYYPGTKLALTEYSYGGENDISGGTAMTDLGLGKNDVYMANWKLK 421
Qy 428 DGANNYSAAKLYRNDYGNKATFGDISVNAQTSDIVNSSVHASVTDASYKELHLIVMKN 487
Db 422 DN-TDYSAAKLYRNDYGNKATFGDISVNAQTSDIVNSSVHASVTDASYKELHLIVMKN 480
Qy 488 SMDSAFADQLDSEITYSSGKLGWDFGDKNSQIKAVAPITQISGNFTYVPLTAYHIV 547
Db 481 NFDDPINALFQSGDKTYTSGRWGFGDQSGDITEQAATINNNQFTTLPPLSAIYHIV 540
Qy 548 LTADNDTPV-----PVEDPEGFTLR----- 568
Db 541 LKADSTEVNLDVYVQKDRNATDNLQIKPHENTONKGTSPV-DLSLTEREYFTKDS 599
Qy 569 -----AEAGDKGVLSWDASSGV-----VGVSVQRATY-ENGFFAAVASNLVE 610
Db 600 SAAMNGWIDWAKLGSNIQISFGNHGADSDTYAELGFSGAGSIAEGGSGEIQLRMSK 659
Qy 611 TSYTDTNVTNGTSY 624
Db 660 ADWSNFNEANDYSF 673

```

RESULT 2

US-08-361-920-29

Sequence 29, Application US/08361920

Patent No. 5457046

GENERAL INFORMATION:

APPLICANT: Woelldike, Helle F.

APPLICANT: Hagen, Frederick

APPLICANT: Hjort, Carsten M.

APPLICANT: Sven, Hastrup

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

TITLE OF INVENTION: or Hemicellulose

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5457046 No. 5457046disk of No. 5457046th America, Inc.

STREET: 405 Lexington Avenue, 62nd Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-361-920-29

```

Query Match 45.9%; Score 1612; DB 1; Length 617;

Best Local Similarity 57.6%; Pred. No. 8e-137;

Matches 307; Conservative 62; Mismatches 150; Indels 14; Gaps 4;

```

Qy 45 VTIKVDTSKDRKPIPIYIGTQDIAAARLGGNMTGYNNWNNMNSAGSDQSS 104
Db 28 VFTTINTOSERAAISPIYIGTQDLSGTENSSRRLLGGNRLTYNNWNNASAGHDWJHY 87
Qy 105 SDFLCNGGLTKACEKPGAVTTSFHQDSILKLGAYSLVTLPMAGYVAKDNGSVQES 164
Db 88 SDFLCNGGVDPDTCDFGAVVTAFAHDKSLGAYSVITLQMGAYVSRDKNGPVDESET 147
Qy 165 APSARNQVNNAKNAPFQLOPDLNDQNVYADEFVNFVKKYGAASAKAGVKGAYALDNEPA 224
Db 148 APSRWDKVEFAKNAPFSLQDPLNDQVYMDVEVNFVKNRYGNASTSTGKAYSILDNEPA 207
Qy 225 LWSHTPHRTHGEKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXI--GT 282
Db 208 LWSETHPRHPEQLQAAELVAKSIDLSKAVKNVDPHAEIFGPAFYGLYSLQDAPDW 267
Qy 283 SVKGNYSWFDVYDQLQRLNSQAEGRLLDVFVHHYPEAMGGGIRIT-NEVCNDETKKA 341
Db 268 SLQGNYSWFDVYDQLQRLNSQAEGRLLDVFVHHYPEAMGGGIRIT-NEVCNDETKKA 327
Qy 342 RMQAPRTLWDPTYKEDSWIAQNSAFLLPLRLKQSDVKYYPGTKLALTEYSYGGENDIS 401
Db 328 RVQAPRSLWDPAQYQEDSWIGTWFSYLLPLRLKQSDVKYYPGTKLALTEYSYGGENDIS 387
Qy 402 GGIAATDVLGILGKNDVYMANWKLKDGANNYSAAKLYRNDYGNKATFGDISVNAQTS 461
Db 388 GGIAATDVLGILGKNDVYMANWKLKDGANNYSAAKLYRNDYGNKATFGDISVNAQTS 446
Qy 462 DIVNSSVHASVTDASYKELHLIVMKNMSDAFDAQDLSGETTYSYSSGKINGFDKNSQIK 521
Db 447 DTENSSVYASVTDENSELHLIVLKNKFDPPINATFQLSGDKTYTSGRWGFGDQSGDIT 506
Qy 522 AVAPITQISGNFTYVPLTAYHIVLADNDTPV-----PVEDPE 564
Db 507 EQAATINNNQFTTLPPLSAIYHIVLADNDTPV-----PVEDPE 559

```

RESULT 3

US-08-479-939-29

; Sequence 29, Application US/08479939
; Patent No. 5686593

GENERAL INFORMATION:

APPLICANT: Woeldike, Helle F.

APPLICANT: Hagen, Frederick

APPLICANT: Hjort, Carsten M.

APPLICANT: Sven, Hastrup

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

TITLE OF INVENTION: or Hemicellulose

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 56865930 No. 56865930th America, Inc.

STREET: 405 Lexington Avenue, 62nd Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,939

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361,920

FILING DATE: 22-DEC-1994

APPLICATION NUMBER: US 07/940,860

FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90

FILING DATE: 09-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00124

FILING DATE: 08-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3435.204-US

TELEPHONE: 212-867-0123

TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 617 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-479-939-29

Query Match 45.9%; Score 1612; DB 1; Length 617;

Best Local Similarity 57.6%; Pred. No. 8e-137;

Matches 307; Conservative 62; Mismatches 150; Indels 14; Gaps 4;

QY 45 VTIKVTSKDRPISPIYIGTQDLADENLAARLGGNRMTGYWNNMNSAGSDWQOS 104

DB 28 VTITINTQSERAAISPIYIGTQDLADENLAARLGGNRMTGYWNNMNSAGSDWQOS 87

QY 105 SDNFCNNGGLTKAECEKPGAVTTSFPHDQSLKLGAYSLVTLPMAGYVAKDNGNSVQESQ 164

DB 88 SDFLCNGGVPTDCKFGAVVATPHDKSLENGAYSVTLQMGVTSRDKNGPVDESET 147

QY 165 APSARNQVNVNAPFQLQDLNDQVYAEFVFLVKYGAASAKGVKGYALDNPPA 224

DB 148 APSRWDKVEFAKNAFSLQPLNDQVYAEFVFLVKYGAASAKGVKGYALDNPPA 207

QY 225 LWSHTPHRIGEVKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXI --GT 282

DB 208 LWSHTPHRIGEVKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXI --GT 282

QY 283 SVKGNYSWFVDYLDQMLNSQAEGKRLLDVFDVHWPYEAQGGQRIIT-NEVGNDETCCA 341

DB 268 SLOGNYSWFVDYLDQMLNSQAEGKRLLDVFDVHWPYEAQGGQRIIT-NEVGNDETCCA 327
QY 342 RMQAPRTLWDPTKEDSWIAQWNSAFLLPLRLKQSDVKYYPGTGKLTALTEYSYGGENDIS 401
DB 328 RVQAPRSLWDPAQEDSWIGTWFSSYLLPLKLOSSQTQYTPGTGKLTALTEYSYGGENDIS 387
QY 402 GGIAWTDVLGILGKNDVYMANWKLDGANNYSAAVKLYRNDGKNATGDISVNAQTS 461
DB 388 GGIAWTDVLGILGKNDVYMANWKLDGANNYSAAVKLYRNDGKNATGDISVNAQTS 446
QY 462 DIVNSSVHASVTDASYKELHLIVMKNKSMDSAFDAQFOLSGETYSYSSGKINGFDKNSSQIK 521
DB 447 DTENSSVAVSVTDEENSELHLIVLKNKFNDDPINATFOLSGDKTYSYSSGKINGFDQSGDIT 506
QY 522 AVAPITOISGNRFTYTPPLTAHYHVLTAANDIPV-----PPVEDPES 564
DB 507 EQAAITNINNQQFTYTLPLSAHYHVLKADSTPEVISEIPSSSTSSPVNQPTS 559

RESULT 4

US-08-483-432-29

; Sequence 29, Application US/08483432

; Patent No. 5763254

; GENERAL INFORMATION:

APPLICANT: Woeldike, Helle F.

APPLICANT: Hagen, Frederick

APPLICANT: Hjort, Carsten M.

APPLICANT: Sven, Hastrup

TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

TITLE OF INVENTION: or Hemicellulose

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 57632540 No. 57632540th America, Inc.

STREET: 405 Lexington Avenue, 62nd Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,432

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361,920

FILING DATE:

APPLICATION NUMBER: US 07/940,860

FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90

FILING DATE: 09-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00124

FILING DATE: 08-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728

REFERENCE/DOCKET NUMBER: 3435.204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 617 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-483-432-29

Query Match 45.9%; Score 1612; DB 1; Length 617;
Best Local Similarity 57.6%; Pred. No. 8e-137;
Matches 307; Conservative 62; Mismatches 150; Indels 14; Gaps 4;

QY 45 VTIKVDTSKDRKPSPIYIYGNQDLADENLAARLGNRTGYNNWNNMNSAGSDWQOS 104
DB 28 VTFTINTQSRAAISPNYIYGNQDLSTGTENWSSRLGNRLTYNNWNNASSAGRWLHY 87

QY 105 SDFLCNNGGLTKAECEKPGAVTTFSDHSLKLGAYSLVLPMAGYVYAKGDCNGSQVSEQ 164
DB 88 SDDFLCGNGVDPDTCRPGAVTAFHDKSLGAYSLVLPQMGAYVYSRKNGVDESET 147

QY 165 APSARWNOVNAKNAPFQLOPLDNDNOVYAFEFYVFLVYKYGAASTKAGYKGYALDNEPA 224
DB 148 APSPRWKEVEAKNAPFSLQPLDNDGOVYMDDEVNVLVYRNGNASTGTGKAYSLDNEPA 207

QY 225 LWSHTPHRIHGEKVKYXXXXXXXNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXI--GT 282
DB 208 LWSETHPHRIHGEQLOAAELVAKSIDLSKAVKNVDPHAEIFGPAHYGFCAYLSLQDAPDWP 267

QY 283 SVKGNYSWFVDYLDQMLNSQAEGKRLLDVDFVHWYPEAMGGGIRIT-NEVGNDETKA 341
DB 268 SLOGNYSWFIDYLDQMKNAHTQNGKRLLDVDFVHWYPEAGGGQRIYVFGAGNIDTQKA 327

QY 342 RQAPRLWDTYKEDSWIAOWNSAFLLPLRLKQSDYKYYPGKTLALTEYSYGGENDIS 401
DB 328 RVOAPRLWDPAYQEDSWIGTWFSYLLPLPKLQSSIQTYYPGKTLAITEFSYGGDNHIS 387

QY 402 GGIAWTDVLGILGNVDMYNNYKLDGANNYSAAKLYRNVDGKNATFGDISVNAQTS 461
DB 388 GGIAWTDVLGILGNVDMYNNYKLDGANNYSAAKLYRNVDGKNATFGDISVNAQTS 446

QY 462 DIVNSSVHASVTDASYKELHLIVNKNMSDAFAQDLSETTYSSGKIWGFKNSSQIK 521
DB 447 DTENSSVYASVTEENSELHLIVLKNFDDPINATFOLSGDKTYTSGRVNGFQTSGIT 506

QY 522 AVAITQISGRNFTYVPLTAYHIVLTADNDTPV-----PPVEDPES 564
DB 507 BOAAITNNNOFTYPLPLSAYHIVLKADSTEPVISEIPSSSTSSPVNQPTS 559

RESULT 5
US-09-136-574A-44
; Sequence 44, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1751 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

Query Match 38.0%; Score 1333; DB 4; Length 1751;
Best Local Similarity 48.5%; Pred. No. 6.1e-111;
Matches 262; Conservative 73; Mismatches 179; Indels 26; Gaps 7;

QY 32 PAPTIVVHGOT-----AKVTIKVDTSKDRKPSPIYIYGNQDLADENLAARLGNRT 86
DB 1215 PPTVTVTPPTPAVTPDKISIDTSRGTKSPIYIYGNQDIQGVVH-PARLGGNRLT 1273

QY 87 GYNWNNMNSAGSDWQOSDNLNNGGLTKAECEKPGAVTTFSDHSLKLGAYSLVLP 146
DB 1274 GYNWNNMNSAGSDWYHSSDDYMCYIMGTGDKNVPAAVVSKFHEQSIKQAYSAITLQ 1333

QY 147 MAGYVAKDCNGSQVSEQAPSAWQVYNAKNAPFQLOPLDNDNOVYAFEFYVFLVYKYG 206
DB 1334 MYGYVAKDCNGTVESETAPSPRWAEVFKDGLSLQPDVNDNYVMDEFINYLKNG 1393

QY 207 AASTKAGVAGYALDNEPALWSTHPRHGEKVKYXXXXXXXNXXXXXXXXXXXXX 266
DB 1394 RSSATGIGKYLIDNEPDLWFTTHPRHPQKVTCSSELINKSVELAKVITLDPDAEIRGP 1453

QY 267 XXXXXXXXFKLHXI--GTSVKGNYSWFVDYLDQMLNSQAEGKRLLDVDFVHWYPEAMG 324
DB 1454 ASYGFVGYLTLDQADPNQVGNHRWFLSWYLEQMKKASDSFGKRLLDVLDLHWYPEAQV 1513

QY 325 GGIRITNEVGNDETK-----ARMQAPRLWDTYK-----EDSWIAOWNSAFLLPL 370
DB 1514 GGVRTCFD-GENSTRDVAIRMQAPRLWDTYKTKQKITAGENSWINQWPEYLP 1572

QY 371 LPRLKQSDYKYYPGKTLALTEYSYGGENDISGGIAMDVLGILGNVDMYNNYKLDGKA 430
DB 1573 LPNIRADIDKYYPGKTLAITEFDYGGKDHISGGIALADVLGIFGKYGYVMAARW--GDS 1629

QY 431 NNYSAAKLYRNVDGKNATFGDISVNAQTSIDIVNSSVHASVTDASYKELHLIVNKNMSD 490
DB 1630 GSYAQAAYNYLYNBDGKSGRYGTCVSAETTDVNNPNVYASTEGEDDSTVHILINRYD 1689

QY 491 SAFDAQDLSETTYSSGKIWGFKNSSQIKAVAPITQISGRNFTYVPLTAYHIVLT 550
DB 1690 RKLKAEIKMNNTRVYTGGEIYGFDTSSQIRKMGVLSNQNNTTIEVPNLTVYHIVLT 1749

RESULT 6
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh Levin, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining

Db 345 USANGNAQVSLTNWAVSGATSYTKRATTSGGPTYNVDRGVATSYNTGLTNGTYYY 404
QY 627 KITAKTKAGTSBNSVLKAVPTPDVDPDYEADGTLKGTIV 668
Db 405 VVRASNSAGSSANSQAASA--TPASG-----GASTGNLV 436

RESULT 8
US-07-952-853-22
; Sequence 22, Application US/07952853
; Patent No. 5863783
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuys, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippin, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES
; TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,853
; FILING DATE: 19921125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-952-853-22

Query Match 3.7%; Score 128.5; DB 2; Length 628;
Best Local Similarity 21.3%; Pred. No. 0.0051;
Matches 88; Conservative 53; Mismatches 123; Indels 149; Gaps 22;

QY 356 EDSWIAQNSAFDLPRL-KQSVKYPGKTLALTEYSY----- 394
Db 254 ENRW--KWNETIGDCRPGREGTWTYNTDGLGLHEFYFECWCDLGLVPVGVWDGFALE 311
QY 395 -GGENDISGGTAMT-----DVLG-----ILKND-----VYMANWK 425
Db 312 SGGNTPUTGD-ALTPYIDDLNLELYILGDTSTTYGAWRAAQQEFPNLTWVEIGNEDM 370
QY 426 LKDGANNYSAAKLYRNYDGKATFGDISVNAOTSDIVNSSVHASVTDASVKELHLYM 485

Db 371 LGGCESEY---AERTAFYDAIHAAYPDLLILIASTSEA--DCLPESMPEGSWDYHDYST 425
QY 486 NKMSDAFAQFDLSGETTYSCKI-----W-----GFDKNSOIK--AV 523
Db 426 PDGLVGQFNFDNLNRSVPYFGEYSRWEIDWPNMKGSAEAVFMIGFERNSDVVKMAAY 485
QY 524 APITQ-ISGNREFT-----YVPP-----LTAYHIVLTAADNDTPVPVPEPESFTLRAEAG 572
Db 486 APILQLINSTOWTPDILGYTQSPGDIFLSTSYV-----QEMFS--RNEG 528
QY 573 DGKVDLSWDASSGVGVSVQRTYENGPPAAVSNLVETSYDTNVTNGTSTYYKTKATK 632
Db 529 DTIKEVTSDS-----DFGLYVWASSA-----GDSYVVKL---A 559
QY 633 KAGTSSENVLKAVPTPV-----DGPDRYEAEDGTL-----KGTIVESSGT 673
Db 560 NYGSETQDLTVSIPGTSTGKLVADSDPDAYNSDTQTLVTPSESTVQASNGT 612

RESULT 9
US-08-914-848-22
; Sequence 22, Application US/08914848
; Patent No. 5989887
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuys, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippin, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,848
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,853
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-914-848-22

Query Match 3.7%; Score 128.5; DB 2; Length 628;

Best Local Similarity 21.3%; Pred. No. 0.0051;
Matches 88; Conservative 53; Mismatches 123; Indels 149; Gaps 22;

QY 356 EDSWIAQNSAFLLPLRL-KOSVDKYYPGTKLALTEYSY----- 394
Db 254 ENRW--KWNETIGDLCDRPGREGTWTYYNTDGLGLHEFYWCEDLGLVPLGVWDGFALE 311
QY 395 -GGENDISGIAWT---DVLG---ILGKND----- 394
Db 312 SCGNTPPLGD-ALTPYIDVLNLEBYILGDTSTTYGAWRAANGQBPWNLTWVEIGNEDM 370
QY 426 LKDGANNYSAAKLYRNYDGNKATFGDISVNAOTSDLVNSVHASVTDASYKELHLIVM 485
Db 371 LGGGCESY---AERTAFYDAIHAAYPDILLIATSEA--DCLPESMEGSDVYHDYST 425
QY 486 NKSMSDAFAQDPLSGTYSGGKI-----W-----GPKNSSQIK--AV 523
Db 426 PDGLVGQFNYFDNLNRSYPYFGEYSRWEIDWPNMKGSAEAVFMEIGFERNSDVVKMAAY 485
QY 524 APITO-IGNRFT-----YTVP-----LTAYHIVLTADNDTPVPPVEDPESFTLRABAG 572
Db 486 APQLQLINSTQWTPDLIGYQSPGDFLSTSIYV-----QEMFS--RNRG 528
QY 573 DGKVDLSWDSAGVGVYQVRATYENGPFPAVASNLVETSYDTNVTNGTSYKKITAKT 632
Db 529 DTIKVETSDS-----DFGPLYWVASSA-----GDSYYVKL---A 559
QY 633 KAGTSESNVLKAVPTPV-----DGPDRYEADGTL-----KGTIVESSGT 673
Db 560 NYGSETQTLTVSIPGTSGTKLTVLADSDPDAYNSDTQTLVTPSESTVQASNGT 612

RESULT 10

US-08-386-727-8
; Sequence 8, Application US/08386727
; Patent No. 5792647

GENERAL INFORMATION:

; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASSIER, BONNIE
; APPLICANT: KEYHAN, NEMAT O.
; APPLICANT: CHITLAPU, EDITH
; APPLICANT: ROWE, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/386,727
; FILING DATE:

CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: HOBBS, ANN S.
; REGISTRATION NUMBER: 36,830
; REFERENCE/DOCKET NUMBER: 4130/206916
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 866 amino acids
; TYPE: amino acid

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-386-727-8

Query Match 3.5%; Score 123.5; DB 1; Length 866;
Best Local Similarity 20.1%; Pred. No. 0.024;
Matches 170; Conservative 89; Mismatches 254; Indels 331; Gaps 52;

QY 25 LIIVGG-----ALPA-----PTVVHGTQAKTYTIKVDTS-----KDRKKLISPIYIG 64
Db 11 LLVGLGCLSTALAAAPKPTIGMGET-KFAIIQVDOAATSYNKLVTVHKDGAHSV----- 65
QY 65 TNODLAGDENLAAR-RLGGNRMGTGYNNWNNMAGSDWQSSDNFLCNGG-----LTKAEC 120
Db 66 TNLWSGDVGOTAKVLLDQKEV-----WSGSAASAAG-----TANFKVTGGRYQOMQVALC 115
QY 121 EKPQAVTTSFHDOSLKLGAISLVTLPMAGVYAKDNGSGVQESQAPSRMNQVYNAKAP 180
Db 116 NADGCTLSDKKE-----IVVADTDG-----SHLAP-----LNAP 144
QY 181 FQLQPDLDNDNOVYADEFVFLVKYGAASAKAGVG--YALDNPEA-----LWSHTRP 232
Db 145 LQ-----ENNKPY-----TNKAGVGVYVWGVYGRKFTVDKIPAKNLTHILYGT-PI 194
QY 233 IHGEKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNYSWFV 292
Db 195 CGNGINDSL-----KEISGSE----- 212
QY 293 DYTLDQRLNSQAEGRLLLDVFDVH--WYPEAMGGGIRITNEVGNDETAKA-----R 342
Db 213 -----ALQSCAGREDFKV-SIHDPWAAVOMQG-----NLTADEPKYKNGNLMALK 260
QY 343 MQAPR-----TLWDPY-----KEDSWIA-----QW----- 363
Db 261 KANPNLKILPSVGGWTLSDPFFFSKTRDIFVASMKEYLOTWKFFDGDVDDWEPGQ 320
QY 364 -----NSAFLPLRLKQSV-----KYYP-----GTLALTEY----- 392
Db 321 GANPNLGGPNMGATYVALMKELRAMLEDEAETGRQYELTSALSAGDKIAKVQYQAAQ 380
QY 393 -----SYGGENDISGGIAMTDVLGILGKNDVYMANVWKLKDGANNVSAAYKLYRNY 444
Db 381 YMDYIFLMST-----DFSGAF-----DLKNLAHQTNLY-ASSW---DPATKYTTD--KGVKAL 427
QY 445 DGKNATFGDISVNA-----QTSDIVNSVHASVTD-----ASYKELHLIVM 485
Db 428 LGQGVTPGKVVVGAAMYGRWTVGVNGYQAGNPFTGSATGPIKGTWENGVDYRD---IV 483
QY 486 NKSMSDAFAQDPLSGTETY-----SSGKIWGFKNSS-OIKAVAPITQISGNRFTYVPP 540
Db 484 NNRMGAGWEQGYDETAEPYVFKASTGDLISFDNDRSVKAKGQYVLANGLGLFAWEIDA 543
QY 541 -----LTAYHIVLTADNDTPVPPVEDP-----ESFTLRAEAGDKVDL-----SWDASSGVYG 588
Db 544 DNGDILNAMHEGLGNGDGGTTPPVNKPVPVANAGSDLSDTGPAEVLNLAASHPDPSGVLS 603
QY 589 YSVQRATYENG-----FAAVAS---NLV-ETSYTDTNVTNGTSYKKITA 630
Db 604 YSWKQV---SGPQVSLLDATQAKARVVLDAVSADINLVFELVTTD-----HNLTA 651
QY 631 KTKAGTSESNVLKAVRTP--VDGPDREAEAG-----TLKGTIVESG-----TGF 675
Db 652 KDQVVVYTNK---APQPNLPPVVTVFATASVESGKQVTKATASDPNGDALTYQWLSLAP 708
QY 676 SGTG 679
Db 709 TATG 712

RESULT 11

US-08-600-452A-8

Sequence 8, Application US/08600452A	343	MOAPR	-----TLWDPT	-----KEDSWIA	-----QW	363
Patent No. 5985644	261	KANPNLKILPSVGGWTLSDPFYFSDTKTRDTFVASWKEYLQTKWFFDGVDDIDWFEFGGQ	320			
GENERAL INFORMATION:						
APPLICANT: ROSEMAN, SAUL						
APPLICANT: BASSLER, BONNIE						
APPLICANT: KEYHANI, NEMAT O.						
APPLICANT: CHITLARO, EDITH						
APPLICANT: ROWE, CHRIS						
APPLICANT: YU, CHARLES						
TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN						
NUMBER OF SEQUENCES: 8						
CORRESPONDENCE ADDRESS:						
ADDRESSEE: FISH & RICHARDSON P.C.						
STREET: 4225 Executive Square, Suite 1400						
CITY: La Jolla						
STATE: CA						
COUNTRY: USA						
ZIP: 92037						
COMPUTER READABLE FORM:						
MEDIUM TYPE: Floppy disk						
COMPUTER: IBM PC compatible						
OPERATING SYSTEM: PC-DOS/MS-DOS						
SOFTWARE: Patent In Release #1.0, Version #1.25						
CURRENT APPLICATION DATA:						
APPLICATION NUMBER: US/08/600,452A						
FILING DATE: 13-FEB-1996						
CLASSIFICATION: 435						
ATTORNEY/AGENT INFORMATION:						
NAME: Haile, Lisa A.						
REGISTRATION NUMBER: 38,347						
REFERENCE/DOCKET NUMBER: 07662/005001						
TELECOMMUNICATION INFORMATION:						
TELEPHONE: (619) 678-5070						
TELEFAX: (619) 678-5099						
TELEX:						
INFORMATION FOR SEQ ID NO: 8:						
SEQUENCE CHARACTERISTICS:						
LENGTH: 866 amino acids						
TYPE: amino acid						
STRANDEDNESS: single						
TOPOLOGY: linear						
MOLECULE TYPE: protein						
US-08-600-452A-8						
Query Match	3.5%;	Score 123.5;	DB 2;	Length 866;		
Best Local Similarity	20.1%;	Pred. No. 0.024;	Indels 331;	Gaps 52;		
Matches 170;	Conservative 89;	Mismatches 254;				
QY	25	IIYGG-----ALPA-----PTVHGQTAKTIVTKVDTS-----KDRKPISPYIYG 64				
DB	11	LLVGLCLSTALAAAPGKPTIGWGET-KFALIQDQAATSYNKKLYTVHKDGAPVSV-----65				
QY	65	TNODLAGDENLAAR-RIGGNRWITGYWNNMNSWAGSDWQSSDNFLCNGG---LTKABC 120				
DB	66	TWNLWSGDVGQTAVKLLDQKEV-----WSGAASAAG-----TANFKVTGKGRYQMQVALC 115				
QY	121	EKPGAVTTSFHDQSLKLGAYSLVTLPMAGVYVAKDNGSVQSEBQAPSAARNQVNAKNAP 180				
DB	116	NADGCTLSDKKE-----IVVADTGG-----SHLAP-----LNAP 144				
QY	181	FQLOPLDNDQVYADEFVNFVKKYGAASTKAGVKG--YALDNEPA-----LWSHTHPR 232				
DB	145	LQ-----ENNKPY-----TNKAGRVGVYVWGVYGRKFTVDKIPAKNLTHILYGET-PI 194				
QY	233	IHGKVKVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXICTSVKGNYSWFV 292				
DB	195	CGGNGINDSL-----KEISGFE-----212				
QY	293	DYILDQMLNSQAEGRKLLDVFVH--WYPEAMGGGIRITNEVGNDETCKA-----R 342				
DB	213	-----ALQRCACAGREDFKV-SIHDPWAAVQMCGQ-----NLTADEPKYKGNFNLMAK 260				

NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 3.5%; Score 123.5; DB 2; Length 1338;
Best Local Similarity 19.8%; Pred. No. 0.048;
Matches 141; Conservative 93; Mismatches 297; Indels 181; Gaps 32;

QY 41 TAKVTIKVDTSKDRKPISPIYIGTGNODLAGDENLAARRLGGNRMWG----- 87
Db 545 TGNITIRQVEGDSR-----VKGVAACKNITFK--GNNITFGSKATTEIKGNVT 594
QY 88 YNNENNMGAGSDWQSSDNF-----LCNNGGLTKAE--CEKPGAVTTFPHDQSLKLGAY 140
Db 595 INKNTNATLGRANFAENKSPNLNAGNVNNGNLTITAGSIINAGNLTVS---KGANLQAI 651
QY 141 SLVTLPMAGYVAKDNGSGVQSEQAPSRWNOVNVNAKNAFQLOPDLNDNOVYADEFVNF 200
Db 652 TNYTFNVAG--SFDNNGASNISARGAKFKDINNNTSS---LNITTSNDSFTY-----RT 700
QY 201 LVKYGAASTKAG-----VKGVALDNEPALWSHTPHRIGHEKVVXXXXXXX 246
Db 701 IIK--GNISNKGDLNIDKSDAEIQIGNISQKEGNL-----TISSDKVNITNQI-- 750
QY 247 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNSVWFVYLDQMLNSQAE 306
Db 751 -----TIKAGVEGGRSDSSE--AENANLTIQTK 776
QY 307 GKRLDVFVHWYPEA-----MGGGIRITN-EVGNDETKKARMAQAPRTLWDPTYKEDSWI 360
Db 777 ELKLAGDLNISGFNKAETAKNGSDLTIGNASGGNADAKVTFDKVK---DSKISTDGHN 833
QY 361 AQWNSAFLPLRLKQSDVYKYPGKTLALTEYSYGGENDISGGIATMDVLGILGKNDVYM 420
Db 834 VTLSN-----EVKTSNGSSNAGNDNSTGIT-----ISAKDVTV 866
QY 421 ANYWKLKGANNYSAAKLYRNYDGK--NATFGDISVNAQTSDIVNSVSHASVTDASYK 478
Db 867 NNNVTSKHTIN--ISAAAAGNVTTKEGTTINATTSVETVTAQNGTIKGNITSONVTVATE 924
QY 479 ELHLIVMKNKSDSAFDAQFDLSGETTYSYSSGKIWGFKNSSOIKAVAPITQISGRNFTYV 538
Db 925 --NLVTTENAVINSGTVNISTKTGDIK---GIESTSGNVNITA-----SGN--TIKV 972
QY 539 PPLTAHYHLVADNDTPVPVDEPESFT-----LRAEAGD--GKVDLSWADSSGVVGSV 591
Db 973 SNITQDVTVTADAGALTTTAGTSISATGTGNANITTKTGDIKGV-----ESSSGSVTLVA 1028
QY 592 QRAVE-----NGPFAVASNLVETSYTDVNTNGTSYVYKITAKTAGTSSES-----N 640
Db 1029 TGAATLAVNISGNTVTITADSGKLTSTVGSTINGTN---SVTTSSQSGDIEGTISGNTVN 1085
QY 641 VLKAVPRPVDGPRYAEADGTLKCTIVESGCTGFGTGYTNFHNAGDSLIT 692
Db 1086 VTATGDLTIGNSAKVEAKNGA--ATLTAESGKLTITQTGSSITSSNGQTTLT 1135

RESULT 13
US-08-719-641-9
; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-9

Query Match 3.5%; Score 123.5; DB 4; Length 1338;
Best Local Similarity 19.8%; Pred. No. 0.048;
Matches 141; Conservative 93; Mismatches 297; Indels 181; Gaps 32;

QY 41 TAKVTIKVDTSKDRKPISPIYIGTGNODLAGDENLAARRLGGNRMWG----- 87
Db 545 TGNITIRQVEGDSR-----VKGVAACKNITFK--GNNITFGSKATTEIKGNVT 594
QY 88 YNNENNMGAGSDWQSSDNF-----LCNNGGLTKAE--CEKPGAVTTFPHDQSLKLGAY 140
Db 595 INKNTNATLGRANFAENKSPNLNAGNVNNGNLTITAGSIINAGNLTVS---KGANLQAI 651
QY 141 SLVTLPMAGYVAKDNGSGVQSEQAPSRWNOVNVNAKNAFQLOPDLNDNOVYADEFVNF 200
Db 652 TNYTFNVAG--SFDNNGASNISARGAKFKDINNNTSS---LNITTSNDSFTY-----RT 700
QY 201 LVKYGAASTKAG-----VKGVALDNEPALWSHTPHRIGHEKVVXXXXXXX 246
Db 701 IIK--GNISNKGDLNIDKSDAEIQIGNISQKEGNL-----TISSDKVNITNQI-- 750
QY 247 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNSVWFVYLDQMLNSQAE 306
Db 751 -----TIKAGVEGGRSDSSE--AENANLTIQTK 776
QY 307 GKRLDVFVHWYPEA-----MGGGIRITN-EVGNDETKKARMAQAPRTLWDPTYKEDSWI 360
Db 777 ELKLAGDLNISGFNKAETAKNGSDLTIGNASGGNADAKVTFDKVK---DSKISTDGHN 833

QY 361 AOWNSAFLPLRLKQSVKYPGKTLALTEYSYGGENDISGGIAMPDVLGILGKNDVYM 420
Db 834 VTLNS-----EYKTSNGSSNAGNDNSTGLT-----ISAKDVTV 866
QY 421 ANYWKLKDGANNYSAAKLYRNDGK--NATFGDISVNAQTSDIVNSSVHASVTDASYK 478
Db 867 NNNVTSKHTIN--ISAAAGNVTTKEGTTINATTSVEVTAQNGTIKGNITSQNVTVTATE 924
QY 479 ELHLIVMKNKSDSAFDAQFDLSGETTYSKGIKFGDKNSQIKAVAPITOISGNRRTYTV 538
Db 925 --NLVTTENAVINATSGTVNISTRTGDIK---GIESTSGNVNITA-----SGN--TLKV 972
QY 539 PPLTAYHIVLTADNDTPVPPVEDPESET-----LRAEAGD--GKVDLSWDSSGVVGYSV 591
Db 973 SNITGQDVTVDAGALTTTAGTISATGNANITTKTGDKNV-----ESSSGSYTLVA 1028
QY 592 QRATYE-----NGPFAAVASNLVETSYDTNVTNGTSSYYKTKAGTSES-----N 640
Db 1029 TGATLAVGNISGNTVTITADSGKLTSTVGSTINGTN---SVTSSQSGDIEGTISGNTVN 1085
QY 641 VLKAVPRTPVDGPDYRAEDCTLKGTVESGSGTGSGTGYVTFNHNAGDSLT 692
Db 1086 VTASTGDLTIGNSAKVEAKNGA--ATLTAESGKLTQTGSSITSSNGQTTLT 1135

RESULT 14
US-08-617-697-9
; Sequence 9, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-9

Query Match 3.5%; Score 123.5; DB 2; Length 1599;
Best Local Similarity 19.8%; Pred. No. 0.064;
Matches 141; Conservative 93; Mismatches 297; Indels 181; Gaps 32;

QY 41 TAKVTIKVDTSKDKRPISPIYICYNQDLADENLAARLGGNRMTG----- 87
Db 805 TGNITIROVEGTDSR-----VNGKVAAKKNTFK--GGNITFGSOKATEIKGNVT 854
QY 88 YWENNMSNAGSDWOSSDNF-----LCNNGGLTKAE--CEKPGAVTTSFHDOSLKLGA 140
Db 855 INKNTNATLURGANFAENKSPNLINAGNVINNGMLTTAGSIINAGNITVS---KGANLQAI 911
QY 141 SLVTLPMAGYVAKDGVSEQAPASRWNVVNAKNAPFQLPDLNDNOVYADEEYVNF 200
Db 912 TNYTFNVAG--SFDNNGASNISARGAKFKDINNNTSS---LNITTSNDDTY-----RT 960
QY 201 LVKYGAAATKAG-----VKGALDNEPALWTHPRIHGEKVKYXXXXXXX 246
Db 961 IIK--GNISNKSGLNIIIDKSDAEIQIGNTISQEGNL-----TISDKYVNIITNQI-- 1010
QY 247 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNYSWFVDYVLDOMRLNSOAE 306
Db 1011 -----TIKAGVEGGRSDSSE--AENANLTIQTK 1036
QY 307 GKRLLDVDFVHWYPEA-----MGGGIRITN--EVGNDETCKKARMQAPRTLWDPTYKEDSWI 360
Db 1037 ELKLAGDLNISGFNKAETAKNGSDLTIGNASGGNADAKKVTDFKVK---DSKISTDGHN 1093
QY 361 AQWNSAFLPLRLKQSVKYPGKTLALTEYSYGGENDISGGIAMPDVLGILGKNDVYM 420
Db 1094 VTLNS-----EYKTSNGSSNAGNDNSTGLT-----ISAKDVTV 1126
QY 421 ANYWKLKDGANNYSAAKLYRNDGK--NATFGDISVNAQTSDIVNSSVHASVTDASYK 478
Db 1127 NNNVTSKHTIN--ISAAAGNVTTKEGTTINATTSVEVTAQNGTIKGNITSQNVTVTATE 1184
QY 479 ELHLIVMKNKSDSAFDAQFDLSGETTYSKGIKFGDKNSQIKAVAPITOISGNRRTYTV 538
Db 1185 --NLVTTENAVINATSGTVNISTRTGDIK---GIESTSGNVNITA-----SGN--TLKV 1232
QY 539 PPLTAYHIVLTADNDTPVPPVEDPESET-----LRAEAGD--GKVDLSWDSSGVVGYSV 591
Db 1233 SNITGQDVTVDAGALTTTAGTISATGNANITTKTGDKNV-----ESSSGSYTLVA 1288
QY 592 QRATYE-----NGPFAAVASNLVETSYDTNVTNGTSSYYKTKAGTSES-----N 640
Db 1289 TGATLAVGNISGNTVTITADSGKLTSTVGSTINGTN---SVTSSQSGDIEGTISGNTVN 1345
QY 641 VLKAVPRTPVDGPDYRAEDCTLKGTVESGSGTGSGTGYVTFNHNAGDSLT 692
Db 1346 VTASTGDLTIGNSAKVEAKNGA--ATLTAESGKLTQTGSSITSSNGQTTLT 1395

RESULT 15
US-09-540-245A-16
; Sequence 16, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 16
; LENGTH: 1381
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-16

Query Match 3.4%; Score 120.5; DB 4; Length 1381;
Best Local Similarity 20.6%; Pred. No. 0.095;
Matches 96; Conservative 66; Mismatches 168; Indels 137; Gaps 25;
QY 316 VHWYPAMGGGIRITNEVGNDETKKARM---QAPRTLMDPT-----YKEDSWIAQWNS 365
DB 192 VHVRRPLIRGPNQTA VGVSSVVFQCRIGGDPDPVLRRTASGGMPLRKFSLHSASG 251
QY 366 AFLPLLPRLKOSVDKYPCTKLALTEYSYGGENDISGGTAMTDVLGI----- 412
DB 252 RVHVLDRSLKLDV---TLEDMEYTCENAV--GGITATGILTVHAPPKFFVIRPKNQ 306
QY 413 ---LGKNDVY--MAN-----YWLKDKANNVSAAYKLYRNYDGKNATGDISVNAQT 460
DB 307 LVEIGDEVLFECQANGHPRTLYWSV-EGNSSLLLPGR-----DGRM-----EVLTPEG 356
QY 461 SDIVNSVHASVTDASYKELHLIVNKK--SMDS-----AFDAQFDLSGETTSSGKIWGF 514
DB 357 RSVL--SIARFARSDGKVVTCNALNAVGSVSRVWVSDTQFELP-PPILIEQGPV--- 409
QY 515 KNSSQIKAV-----APITQIS-----GNRFTYVVPPLT----- 542
DB 410 NOTLPVKSIWLPCTRLGTGPVQVSWYLDGIPIDVQEHERRNLSDAGALTISDLQRHEDE 469
QY 543 -AYHIVLTADN-----DTPVPP-----VEDPESEFTL-----RAEAGDGKVDL 578
DB 470 GLYTCVASNRKSSWSGYLRDLDTPTNPNIKFFRAPELSYPPGPKQPMVEKGENSVTL 529
QY 579 SWDAS-----SGVVGYSVQR-ATYENGPPAAVASNLVETSYDTNVTNGTSTYYKITAKT 632
DB 530 SWTRSNKVGSSLVGYVIEFMGKNETDGVAVGTRVQNTTFTGTGLLPVNVYFFLIRAE 589
QY 633 KAGTSES NVLKAVPRTPV-----DGPDRYEAEDGTLKGTIVESS 671
DB 590 SHGLS-----LPSMSEPTVGTGRYFNSGLDLSEARASLLSGDVVELS 632

Search completed: July 13, 2002, 00:49:45
Job time: 1776 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:48:58 ; Search time 59 Seconds
(without alignments)
1131.900 Million cell updates/sec

Title: US-09-784-554b-6

Perfect score: 3509

Sequence: 1 MKAKNSSNLSKRSKWLPPV.....SGTGYVTFHFNAGDSLTMTI 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1617.5	46.1	700	2 B41897	cellulase (EC 3.2.
2	1374	39.2	606	2 H97012	hypothetical prote
3	1278.5	36.4	1331	2 A48954	mannan endo-1,4-be
4	243	6.9	1203	2 S27545	pullulanase - Ther
5	232.5	6.6	1475	2 A44765	alpha-amylase (EC
6	223	6.4	1481	2 S28669	amylolopulanase p
7	161.5	4.6	1090	2 AG1749	glycosidase homolo
8	161	4.6	1684	2 S10789	amylase A-180 - al
9	157	4.5	1042	2 A97209	sporid-like domain
10	154.5	4.4	1091	2 AF1380	glycosidase homolo
11	148	4.2	2032	2 J39917	hypothetical prote
12	146.5	4.2	1282	2 JC4393	microbial collagen
13	143	4.1	872	2 S49541	cellulase - Cellul
14	141.5	4.0	1433	2 T30261	chitinase (EC 3.2.
15	140.5	4.0	1090	2 S59077	cellulose 1,4-beta
16	136	3.9	2167	2 AF1489	cell wall-associat
17	135.5	3.9	699	2 A38368	chitinase (EC 3.2.
18	135	3.8	976	2 A42466	alpha-amylase (EC
19	134.5	3.8	1045	2 A39199	endoglucanase B (E
20	134.5	3.8	26926	1 I38344	titin, cardiac mus
21	133.5	3.8	6831	2 A88852	protein unc-22 (im
22	133.5	3.8	6839	2 S57242	twitchin (similar
23	133.5	3.8	7160	2 T27935	hypothetical prote
24	132	3.8	583	2 S57721	cspB protein - Clo
25	131	3.7	157	2 T44794	hypothetical prote
26	130	3.7	6805	2 S20901	titin - rabbit (fr
27	128.5	3.7	824	2 F87290	TonB-dependent rec
28	128.5	3.7	1441	2 B86807	hypothetical prote
29	128.5	3.7	1578	2 AD1512	peptidoglycan boun

30 127 3.6 488 2 A41961 chitinase (EC 3.2.
31 127 3.6 817 2 A48721 titin, muscle - ch
32 127 3.6 2348 2 AD1841 hypothetical prote
33 125.5 3.6 2044 2 AB1180 probable peptidogl
34 124.5 3.5 747 2 D47093 cellulase (EC 3.2.
35 124.5 3.5 1063 2 D83789 hypothetical prote
36 124.5 3.5 1386 2 AC1533 surface protein (L
37 124 3.5 967 2 S66852 hypothetical prote
38 124 3.5 1268 2 AB0204 conserved hypothet
39 123 3.5 1959 2 AG1085 hypothetical prote
40 122.5 3.5 478 2 AC1383 chitinase and chit
41 122 3.5 821 2 AD1148 probable secreted
42 122 3.5 980 2 H50681 probable flagellin
43 122 3.5 980 2 D85532 probable structura
44 121.5 3.5 2183 2 T37218 hypothetical prote
45 121 3.4 440 2 I50213 protein-tyrosine-p

ALIGNMENTS

RESULT 1

B41897

cellulase (EC 3.2.1.4) - Bacillus lautus

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus lautus

C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999

C:Accession: B41897; S27499

R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A:Title: cels from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-

A:Reference number: A41897; MUID:92276330

A:Accession: B41897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-700 <HAN>

A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663

A:Experimental source: PL236

A:Note: sequence extracted from NCBI backbone (NCBIP:104605)

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans, such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 46.1%; Score 1617.5; DB 2; Length 700;
Best Local Similarity 48.7%; Pred. No. 3.9e-105;
Matches 328; Conservative 83; Mismatches 200; Indels 63; Gaps 9;

Qy 11 SKRSKWLPPVMWACTIIVGALPAPTIVHGTAKVTIKVDTSKDRKPISPYIYGTNODLA 70

Db 3 TRQRKLFVSAALAVSLTMTVMPASVNA-AASDVTFITQSERAAISPNIYGTNODLS 61

Qy 71 GDENLAARLGGNRMTGYNNWNNMNSAGSDWQSSDNFLCNGGLTKAECEKPGAVTTSF 130

Db 62 GTENSSRRLLGNRLTYGNWENNASSAGRWLHYSDDFLCNGGVPDTCDKPGAVTAF 121

Qy 131 HDQSLKLGAYSILVTLPMAGYVAKDNGSVQSEQAPSARWQVNNAKNAPQLQPDLDND 190

Db 122 HDKSLENGAYSILVTLQMGAGYVRDKNQGVDESEAPSPRWKVFNAFSLQPHLNDG 181

Qy 191 QVYADEVFNVLVKKYGAATSKAGVKGVKALDNEPALWHSHTPHRIGEKVYKXXXXXXX 250

Db 182 QVYDEEYVFLVRYGNASTGTGKAYSLDNEPALWSETHPRIHPEQLQAAELVAKSIDL 241

Qy 251 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXI--GTSKGVNYSWFVDYDYLQMLNSAQEG 308

Db 242 SKAVKNVDPAEIEFGPALYGFCAVLSLQDAPGWSPSLQGNYSWFIDYDYLQMLNSAQEG 301

Qy 309 RLLDVFVHWTYPEAMGGGIRIT-NEVGNDETCKKARMQAPRTLWDPTTKEDSWIAQWNSAF 367

Db 302 RLLDVLVDVHWTYPEAMGGGQRIYVFGAGNIDTQKARVQAPRSLMDPAYQEDSWIGTWFSY 361

QY 368 LPLPLRLKQSDVKYYPGKTLALTEYSYGGENDISGGIAMTDVLGILGKNDVYMYW 427
Db 362 LPLPLKQSDVYYPGKTLALTESSYGGNDHISGGIATADALGIFGKYGVYAANYWTE 421
QY 428 DCANNYSAAYKLYRNDYDGNKATFGDISVNAQTSDDIVNSSVHASVTDASKEHLIVMK 487
Db 422 DN-TDYTSAAYKLYRNDYDGNKATFGDISVNAQTSDDIVNSSVHASVTDASKEHLIVLNK 480
QY 488 SMDSAFADQFDLSETTYSYSGKIDGPRDKNSOIKAVAPITQISGNRETYVTPPLTAYHIV 547
Db 481 NFDDPINATFOLSGDKTYTSGRWGFDQTSDETEQAATININNNFTYTPPLSAYHIV 540
QY 548 LTADNDTPV-----PPVEDPESEFLR----- 568
Db 541 LKADSTEPVNDLVQYKDGDRNATDNQIKPHFNQKGTSPV-DLSSLTLYFTKDS 599
QY 569 -----AEADGKVDLSWDASSGV-----VGYSVQRATY-ENGPFPAVASNLVE 610
Db 600 SAAMNGWIDWAKLGGSNIOISFGHNHNGADSDTYAELGFSGAGSIABGGSGEIQLRMSK 659
QY 611 TSYTDNTVNTGTSY 624
Db 660 ADWSNFEANDYSF 673

RESULT 2

H97012
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97012
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-606 <R>
A:Cross-references: GB:AE001437; PIDN:AAK78891.1; PID:gl5023815; GSPDB:GNO0168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0915

Query Match 39.2%; Score 1374; DB 2; Length 606;
Best Local Similarity 45.5%; Pred. No. 3.4e-88;
Matches 282; Conservative 84; Mismatches 216; Indels 38; Gaps 11;
QY 12 KRSKWLPPVMACTIIVGGALPAPT---VHGTQAKTVTIKVDTSKDRKPISPIYIGTNQD 68
Db 2 KRNKIL--VSALAIATFSSMLTPVSGLKVFADTA-DVNVNIDTNAEKQAIISPIYIGTNQD 58
QY 69 LAGDENLAARLGNRMRTGYNWNMNSAGSDWQSSDNFLCNNGGLTKAECEKPGAVTT 128
Db 59 FS-NAKVTAARRIGNRSTGYNWNNDNAGTDWKNESDNTWLTLDYDPVPEKYNEPASVYT 117
QY 129 SFHQSLKGLA-YSLVTLPMAGYVAKDNGSVQSEQASARWNOVYNNAKNAFPQLQPD 187
Db 118 AFHDKSLAMGVPYSLVTLQAGGYVAAQDGLPLANTDVAPSKWKVFEFNKNGPLSLTPDT 177
QY 188 NDNOVYADEFNFLVKYGAASTKAGYKGYALDNEPALWSHTPHRIGERKVV--XXXXXXX 245
Db 178 TDGSYMDDEFVNLVKNYKGSAGSKGKGYSLDNEPSLWSTPHLPHDKTKCEVLDKD 237
QY 246 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVGNYSWFVDYDLDQMLNSQA 305
Db 238 TQLAQVVKKIDPAATETGPALEGFSAFNDENSSPDWSDSVKGNQWETDYDLDNNKKNSDA 297
QY 306 EGKRLDVEDVHWYPEAMGGGIRI-TNEVGNDETKKARMOAPRTLWDPTYKEDSWTAOWN 364
Db 298 AGKRLDLDLHWYPEAKGGGQRTTSDTSNVDCNKAQAPRLSDMTSTYTEDSWIGQWC 357

QY 365 SAFPLPLRLKQSDVKYYPGKTLALTEYSYGGENDISGGIAMTDVLGILGKNDVYMYW 424
Db 358 KWGLPLPKVKSIDKYYPGKTLSEYNGGEDIHSGIAQDALGVFGKYGVYFATYW 417
QY 425 KLDGANNYVSAAYKLYRNDYDGNKATFGDISVNAQTSDDIVNSSVHASVTDASKEHLIV 484
Db 418 ECNSOKNNYVQSAFNLNNDYDGNNSKYGDTDKVCTSDINNSSTYASVTSNDGNKMDIIV 477
QY 485 MNKSMDSAFADQFDLSETTYSYSGKIDGPRDKNSOIKAVAPITQISGNRETYVTPPLTAY 544
Db 478 MNKNTDTSINFNFNSSNNKNTYSGQWGFDSNSNITRRDDVSSISGNKFTYKIPALTA 537
QY 545 HIVLTADNDTPVPPVEDPESEFLRAEAGDKVDLSWDASSGVVGYSVQRATYENGPFPAV 604
Db 538 HIVLIT-----AQKSSVKGDVNGD---GVNV--GRDLMLKRYLAG 573
QY 605 ASNLVETSYTDN---VTNG 621
Db 574 STSNIDLNAADLNNDGAVNG 593

RESULT 3

A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocoellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocoellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocoellum saccharolyticum" is part of a multidoma
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:gl44290; PIDN:AAA71887.1; PID:gl44291
A:Note: sequence extracted from NCBI backbone (NCBIN:121576, NCPIP:121577)
R:Luehthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPROHOHQ' <LUE>
A:Cross-references: EMBL:M36063; NID:gl44292; PIDN:AAA72861.1; PID:gl44294
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 36.4%; Score 1278.5; DB 2; Length 1331;
Best Local Similarity 45.7%; Pred. No. 5.1e-81;
Matches 249; Conservative 85; Mismatches 190; Indels 21; Gaps 6;
QY 32 PAPTVMHGTAKTVTIKVDTSKDRKPISPIYIGTNQDLAGDENLAARLGNRMRTGYNWE 91
Db 777 PLPTI--SPSPSVVEITINTNAGRTQISPIYIGANQDIEGVVH-SARRLGNRLTGYNWE 833
QY 92 NNMNSAGSDWQSSDNFLCNNGGLTKAECEKPGAVTTSFHDOSLKLKAGYSLVTLPMAGYV 151
Db 834 NFNNSAGNDWYHSSDDYLCWSMGSIGEDAKVPAAVVSKFHEYSLKNAYSAVTLQWAGYV 893
QY 152 AKDNGSVQSEQASARWNOVYNNAKNAFPQLQPDNDNOVYADEFNFLVKYGAASPK 211
Db 894 SKDNYGTVSENETAPSNRAEYKFKKADPLSLDNDNFVYMDDEFINLYNKYMASSP 953
QY 212 AGVKGYALDNEPALWSHTPHRIGERKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXX 271
Db 954 TGKGYILDNEPDLWASTPHRIHPNKTCKELIERSVELAKVIKTLDPSEAVFGYASYGF 1013

QY 272 XXFXKLHXI -GTSVKNYSWFVDDYLLDOMRLNSOAEKRLLDVDFVHWYPAMGGGIRI 329
Db 1014 MGYISLQADPDNVOVKEHRWISWLEOMKRASDSFGKRLLDVLDLHWYPEARNGNIRV 1073
QY 330 TNEVGNDTKK---ARMOAPRTLWDPTYK-----EDSWIAQWNSAFPLPLRLKQ 376
Db 1074 CDFGENDTSKEVVIARMOAPRTLWDPTYKTSVKGOITAGENSWINQWFSYDILPIPNVKA 1133
QY 377 SVDKYYPGPKLALTEYSYGGENDISGGTAMTDVILGKNDVYMANWKLDGANNYVSA 436
Db 1134 DIEKYPGPKLALTEYSYGGENDISGGTAMTDVILGKNDVYMANWKLDGANNYVSA 436
QY 437 AKLYRNYDKGNATGDISVNAQTSDIVNSVSHASVTDASYKELHLIVNKSMDSAFDAQ 496
Db 1191 ANYIYLYDGKSKYGNVNSANTSDVENMPYASINGQDDSELHILINRNYDKLOVK 1250
QY 497 FDLGSETYSKKGKNGFKNKSQIKAVAPITQISGNRRYTYVPLTAYHIVLTADNDPPV 556
Db 1251 INITSPKYTKAEIYGFDSNPSPEYKMGNIENSVETLEVPKNGVSHSTILDNFVSI 1310
QY 557 PPVED 561
Db 1311 KIIQN 1315

RESULT 4
S27545
pullulanase - Thermoanaerobacterium thermosulfurigenes
C:Species: Thermoanaerobacterium thermosulfurigenes
C:Date: 09-Jun-1994 #sequence_revision 26-May-1995 #text_change 26-May-1995
C:Accession: S27545
R:Burchhardt, G.; Haecckel, K.; Spreinat, A.; Antranikian, G.; Bahl, H.
submitted to the EMBL Data Library, March 1992
A:Description: Nucleotide sequence of the pullulanase gene from Clostridium
A:Reference number: S27544
A:Accession: S27545
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1203 <BUR>
A:Cross-references: EMBL:M57692
A>Note: the source is given as Clostridium thermosulfurogenes

Query Match 6.9%; Score 243; DB 2; Length 1203;
Best Local Similarity 20.8%; Pred. No. 1.2e-08;
Matches 138; Conservative 86; Mismatches 228; Indels 210; Gaps 30;

QY 62 IYGTNQDLADENLAARLGNRTGYWNNMNSA-----GSDWQSSDNFLCN 111
Db 497 MEGTTQD-----FEKLMSDAHAAGIKIILDGVFNHTSDDSIYFN 535
QY 112 NGGLTKAECEKPGAVTTSFHDQSLKLGAYSLVTLPMAGYVAKDNGSVQSEQAAPSARN 171
Db 536 RVG-----KYPG-----LGAYQ-----AWKEGNSL-----SPYGDW- 562
QY 172 QVNNAKNA-----PFQQLPDLNDNOVYADEFVFLVKYGAAS-----TRAGV 214
Db 563 YTINSDGTYECWGWGDSLPVTKSLNGSEYNTVSWANFIINDENALISKYWLNPDLNDGA 622
QY 215 KYALD--NEPA--LWSHTHPRIHGEKVKVXXXXXXXXXXXXXXXXXXXXXXXXXXXX 270
Db 623 DQWRULDVENEVAHDPWTHFRNAINTVKFEAPMAENMGDA----- 662
QY 271 XXXXFKLHXIGTSVKRG--NYSW--FVDYILDOM-----RLNSOAEKRLLDVDFV 316
Db 663 -----SLDLLGDSFNSVMNYQFRNDIDFLIGQSFDCNGQHNPDIAKLDQRLMSIYER 717
QY 317 HWYPAMGGGIRITNEVGNDETKKARMOAPRTLWDPTYKEDSWIAQWNSAFPLPLRLKQ 376
Db 718 YPLAFYS-----TMNLGSHDTRMLLTVEFGYNSADPNNSDA-AKQLAQKIKLATILQM 772
QY 377 SVDKYYPGPKLALTEYSYGGENDISGGTAMTDVILGI-LGKNDVYMANWK----- 425

Db 773 G-----YPG-----MADIYGDDEAGVSGGKDPDRRTFPWGNEDTTLQDFFKNISIRNNQ 824
QY 426 -LKDCANNYSAAAYKLY-----RNYDGKNATFGD-----ISVNAOTSD--IVNSSVHA 470
Db 825 VLKGTDLLETLYAQNDDVYAGRIIRINGKDA-FGTSYPDPSAAIVAINRSKSKQIAIDTKF 883
QY 471 SVTDAASYKELHLIVNKSMDSAFDAQFDLSETTYSYSGKIGWFDKNSQIKAVAPITQIS 530
Db 884 LRDCWTFKDL-----INNVSYS-----IS 903
QY 531 GNRETYVPLTAYHIVLTADNDPPVPPVEDPESFTLRAEAGDKVDLSWDASSGVVGY 590
Db 904 NGQIVIDVPMAG--VMLISDDGQDLTAPOAPSNNVV--TSGNGKVDLSLQSDGATGYN 959
QY 591 VQRATYENGPFPAAVASNLVETSYTDTNVTGTSYIYKITAKTAGTSE--SNVLKAVPRT 648
Db 960 IYRSSVEGLYEKIASVNTETTFEDANVTNGLKYVYSAIDELGNSGSIINDAVAPAY 1019
QY 649 PV 650
Db 1020 PI 1021

RESULT 5

A44765
alpha-amylase (EC 3.2.1.1) / alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41)
C:Species: Thermoanaerobacter thermohydrosulfuricus
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 15-Oct-1999
C:Accession: A44765
J. Gen. Microbiol. 136, 447-454, 1990
A:Title: Nucleotide sequence of the alpha-amylase-pullulanase gene from Clostridium
A:Reference number: A44765; MUID:90362027
A:Accession: A44765
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <MEL>
A:Cross-references: EMBL:M28471; NID:g144726; PIDN:AAA23205.1; PID:g144727
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: fibronectin type III repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 6.6%; Score 232.5; DB 2; Length 1475;
Best Local Similarity 22.8%; Pred. No. 8.7e-08;
Matches 140; Conservative 69; Mismatches 190; Indels 215; Gaps 33;

QY 132 DQSLKLGAYS--LVTLPMAGYVA-KDNGSVQSEQAAPSARNQVYVNAKNAPFQLQPD-- 186
Db 527 DDSIYFDRYGKYLNTGVLGAYQAWKQDQS-----KSPYGDW-----YEIKPDGT 571
QY 187 -----LNDNOVYADEFVFLVKYGAAS-----TRAGVGYALD-- 220
Db 572 YEGWGWGDSLPVIRINGSEYVKSADFIINPNPNAISKYWLNPDKGNVGDGWLRLVA 631
QY 221 NEPA--LWSHTHPRIHGEKVKVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLH 278
Db 632 NEVAHDFVWHFRGAINTVKPNAPMAENWMDA-----SLD 666
QY 279 XIGTSVKRG--NYSW--FVDYILDOM-----RLNSOAEKRLLDVDFVHWYP-- 320
Db 667 LLGDSFNSVMNYLFRNAVIDFLDKSFDCNGVNVNPDIAKLDQRLMSIYERYPLPVFVS 726
QY 321 --EAMGG--GIRITNEVG-NDETKKARMOAPRTLWDPTYKEDSWIAQWNSAFPLPLRLK 375
Db 727 TMNLGSHDTRMLLTVEFGYNSADENQNSQAARDL-----AVKRLK 766
QY 376 QS--VDKYYPGPKLALTEYSYGGENDISGG-----IAMTDVILGI 414
Db 767 LAAIQLMGYPG-----MPSIYVYGDGAGSGGKDPDRRTFPWGNEDTTLQDFFKNVNI 822

QY 170 WQVNAKNAPOLOPDLNDNOVIAD-----EFVNFLLKYGGAAGTAKGVKGYALDNEPAL 225
 Db 528 W-----FGNGLLLDFTNKDAVNWTSOREYLLLDVGDGFKTGGEMV 570
 QY 226 WSTHPRRIHGEKVKYXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSYK 285
 Db 571 WGRDITTSNGEK-----GQEMR 587
 QY 286 GNY-SMFVDYIYLDOMR-LNSQA-----EGKRLLDVDFVHWPAMGGGIRITNEVNDDET 338
 Db 588 NRIPTYVSYDFPAKINPEAVSFRSGTSGAOKSGIYW-----SGDOT 632
 QY 339 K-----KARMOAPRT-----LWD-----PT-----YKEDSWIAQWNSAFLLPLR 373
 Db 633 STFDFOASQAGLSASTSGVSWAMDAGFTGNYPTAELYKRATAMA-----AFAPIMQF 688
 QY 374 LQSVDKYYPGKTLATEYSYGGENDISGGIAMTDLVGLILKNDVYMAN-----YWK 425
 Db 689 HSEKSDP-----SPSEERSPNNAVARTGDETILPTFKYLYTRMNLPLPYITA 736
 QY 426 LKDGANNVYSAAYKLYRNYD-----GKNATFGD-----ISVNAOTS-----461
 Db 737 AKDTADNGKSMRQAMWDYPEDINTNLDEQYWFQDGLLVAVIYVGGQTEKEVYLPGEW 796
 QY 462 DIVNSSVHASVTDASY-----KELHLIYMN-----KSMDSAPDAQFDL--499
 Db 797 IDIWWGIHPGGETISYADVDTLPVFAKAGALIPMNMTDGYQLGQNGVNDLKAVDNLTF 856
 QY 500 ----SGETYSCKINGFKNSQIKAVAPITQISGNRTYTPPLTAYHIVLTADNDTP 555
 Db 857 RYVPSGNSYS-----FYDVGNGEMRDISVSEDFANEKVTYNLPAM-----ADETM 904
 QY 556 VPPVEDPESFTLR-AE-AGDGKVDLSWASSGVGVSVQVADYENGPFPAASVNLVETSY 613
 Db 905 QVFESEPTSVTVAGAEVAKADTLDAFENATSAYYDVTQNLTY-----IKTAA 952
 QY 614 TDTNVTNGTSYYKIKTAKTSGESNVLKAVPRTPVDGPDYEAEDGTLKGTIVSESGT 673
 Db 953 TD-----TTQAILVGVNHAP-----YEAFFGLHNVSTASDHA 986
 QY 674 GFSGCTGYTNFHNAGDSLWTMTI 695
 Db 987 GYTGTGTFVAGFDEKEAVEFDI 1008

RESULT 8
 S10789
 anylase A-180 - alkaliphilic eubacterium 163-26
 C:Species: alkaliphilic eubacterium 163-26
 C>Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 15-Oct-1999
 C:Accession: S10789
 R:Candussio, A.; Schmid, G.; Boeck, A.
 Eur. J. Biochem. 191, 177-185, 1990
 A>Title: Biochemical and genetic analysis of a maltopentaose-producing amylase from an
 A:Reference number: S10789; MUID:90336627
 A:Accession: S10789
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-1684 <CAN>
 A:Cross-references: EMBL:X53373; NID:948305; PIDN:CAA37453.1; PID:948306

Query Match 4.6%; Score 161; DB 2; Length 1684;
 Best Local Similarity 20.6%; Pred. No. 0.011;
 Matches 146; Conservative 62; Mismatches 256; Indels 244; Gaps 29;

QY 64 GYNQDLADENLAARRLG-----GNRMTGYNNWNNMNSAGSDWQ---QSSDNFLCNN 112
 Db 461 GGQVQFYGDE--TARPLDGGSDPDQGT-R-SSMWNANQNVLHWKLGQFRNHTAIG 517
 QY 113 GGLTAECEKPCAVTTSFHDQSLKLGAYSLVTLPMAGYVAKDNGSVQSEQAPSAKWNQ 172
 Db 518 AGAHQKLSDSPYTFATYESDDI-----VDEVVVAAGQOTT 554

QY 173 VVNAKNAPFOLPDLNDNOVIADFEV--NFLVKYKGAAGTAKGVKGYALDNEPALWSHTH 230
 Db 555 AVTVEGV-----FEDGTVVRDAYTGDETTATK-GTATFTAGTQGIILIENTA-----600
 QY 231 PRIHGEKVKYXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSYKGYNSW 290
 Db 601 ----EPVTNL-----PIVSATPGNSSF 618
 QY 291 FVDYLDOMRINSQAEGKRLLDVDFVHWPAMGG-GIRITNEVNDDETAKRMOA-----345
 Db 619 RTDDITITLNVDRADMGKYTLDSGDPADGLTFMDGEEIVIGADMEDFETATRLIYAENEN 678
 QY 346 -----PRLWDPYTKEDSWIAQWNSAFLLPLRLKOSVDKYYPGT-----385
 Db 679 GIRTSYTYRKVDPDALLEVYFKKP---ADMGTPIH-----YYDTPEBPEV 723
 QY 386 -----KLALTE---YSYGGENDISGGIAMTDLVGL--ILGKNDV-----YMANVWK 425
 Db 724 TWTAPEMTLVEDDWWYVVFENAESANLIFKDSGKQIIPGNEPGFFIDQIGWYDGVKWL 783
 QY 426 LKDGANNVYSAAYKLYRNYDGNKNAFTGDISVNAOTS-----DIVNSSVHASV 472
 Db 784 DSDPFEREKPEPATTPKNNLSVNVNVTETVTFEWDQSDGYVVEYELRDEDDVASTIRTF 843
 QY 473 TDASYKELHLIYMNKMSDMSAFDAQFDLSGETTYSYSGKIWFQDNSSOIKAVAPITQISGN 532
 Db 844 TDE-----DLNPDITY-----TYSVAVVGEQKSA-869
 QY 533 RFTYVPLPPLTAYHIVLTADNDTPVPPVEDPESFTLRAEAGDGKVDLSWASSG-VVGYSV 591
 Db 870 -----PSEALKVTTLLEENDEKPEAPEAPENLRI-ADITDTVTIWNKASNGYVGYEV 921
 QY 592 QRATYENGPFPAASVNLVETSYDTNVTNGTSYYKIKTAKTSGESNVLKAVPRTPVD 651
 Db 922 LRD-----GVVIGETTRTFIDTGLDADRTYTTIVALGDGG-QKSDPSEALEVTQTQ 973
 QY 652 GPD-----RYEAEDGTLKGTIV-----ESSGTGFS 676
 Db 974 KPEGNLVTIYKKGDFPYMHYRPEGG--EWTIVPGIRMBESETAGYS 1019

RESULT 9

A97209
 SpoIID-like domain containing protein, peptidoglycan-binding domain [imported] - Clos
 C:Species: Clostridium acetobutylicum
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: A97209
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: A97209
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1042 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK80460.1; PID:g15025528; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2506

Query Match 4.5%; Score 157; DB 2; Length 1042;
 Best Local Similarity 20.0%; Pred. No. 0.01;
 Matches 159; Conservative 92; Mismatches 283; Indels 260; Gaps 40;

QY 28 GGALP-----APTVVHGTAKTVT-----IKVDTSKDRKPIPIFYIGTNQDLAGE 73
 Db 241 GFALPYLISKPDVINGQTVDNASWQGNKSFNSASIDTTLKSK---GYLLSTDTSFVSIDL 297
 QY 74 NLAARRLGG-----NRMTGYNNWNNMNSAGSDNQSSDNFLCNGGL 115

298 NSITRPSGRVSSVSIYIKDSTGTGTSKAITSDKCRFTFLNLQSSMWYLSYDSV---NGIY 354
116 TKACEKPGAVTTTFHDSQSL---KLGAYSLVTLPM---AGVAKDNGSGVQESQAPSAW 170
355 T-----FTGKGYGHGLMSQIGAYQANLQGNVTGILNFYNNISIVD----- 396
171 NOVYNNAKNAFQLOPDLNDNOVYADEFVFLV---KKYGAASTKAGV---KGYALD---N 221
397 NVIKYAKLSVY---INDSTNNQIYLGQIVNYVSALQDQGYGNLFKFGYIKNGAVLSETDYN 454
222 EPALWSTHPIRHEKGVYXXXXXXXKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIG 281
455 SNNTFSET-PTSSGD-----YQVY 472
282 TSVAGNYSWFYDYLD---QMLNSQA-----EGKRLDVFDP-----VHWYPEAMGG- 325
473 VKIKDKYS---DKTFDDTLSSLNVQAIIPAVSISGTFEDTDSIVSQKSNFNASASGGT 529
326 --GIRITNEVGNDETAKKMQAPRTLWDPTYKEDSWIAQWNSAFPLPLLRKQSVDPKYYP 383
530 GLGYLYKYEISKDGV---VQTTNYSSTPTF-----TVNPNISGVYT 568
384 GT---KLALTEYSYGGENDISGGTAMTDVLGILGKNDVYMYW-----KLKDGANNYV 434
569 STLXVKDTLSNAAVDSKSLNFIYSPSISALNLSKIELLQGGQSTDFAPSITGGSGNYL 628
435 SAAYKLYR-----NYDGKNATFGDISVNAQTSDIVNSVSHASVTDASVKEL 480
629 -KYEYIRNGNLATQDQFASNSYSPTNFTGDSVVKLYVKDALSTSDYDAVTSANEKTY 687
481 -----HLIVNKSMDSAFDAQF---DLISG-----ETYSYSGKIWGF 513
688 AAPALSSPLDSPOMISGSSANFSAQISGSGSNPSYKFAVYNNGLSLVTDLTLSLNSAFS 747
514 DKSQIKAVAPITQISGNRTYTPPL---TAYHIV---LTADNDPVPVPE-----D 561
748 SPSTAGIYQV-----VYVYKRLTDAYSVMGSLTVFNVKISNVASGYMYD 797
562 PESETLRAEAGDGRVLDLSDWASSGVYGVQVATYENGPFPAVASNLVETSYTDITNVYNG 621
798 GKPTLTNTTDDGSTS-----GFNVRYEYKNGALVSSSDYSISPNYFTPTTYG 847
622 TSYKKITAKTKAGTSNSVLKA-----VPRTPVDG---PDYEAEDG---TLKGTIVE 669
848 T---YTKAFYKDLNLSINN-LDAEKDFSIYNKAPLSAALPLNYGMTSGDVTNLQNALIY 903
670 -----SSGTGFSGT 678
904 LGYSSISATGYGT 917

RESULT 10
AF1380
glycosidase homolog lmo2446 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1380
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Meok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1380
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1091 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00524.1; PID:g16411934; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2446

Query Match 4.4%; Score 154.5; DB 2; Length 1091;
Best Local Similarity 18.7%; Pred. No. 0.016;
Matches 148; Conservative 81; Mismatches 270; Indels 291; Gaps 35;
QY 39 GOTAKTIVTIKVTSDKRPISPIYITQDLAGDENLAARRLGGNMTGYNMENN----- 93
Db 355 GDMTNMLDYVIVSGKDQND---IVNNYTDITGKTLPLPKWAFGLWMSANEMDRESYSS 410
QY 94 -MSNAGSD-----WQSSDNFLCNGGLT-----KACEKPGANV 126
Db 411 ALSNAKANDIPATGFVLEQMSDEEYIYINNATYAKNGEAFSYDDFTFNGKWTDPKGM 470
QY 127 TFSHDSOLKGLAYSILVTLV-----NAGYVAKDNGSGVQSEQAPSA 169
Db 471 VDSVHDAGNIVLWQVPVLKDDGTVEQORDNDDEEYIWSQYSADDTGA---PYRVPASQ 527
QY 170 WQVYNNAKNAFQLOPDLNDNOVYAD---EFVNFVLVKYGAASTKAGVKGYALDNEPAL 225
Db 528 W-----FNGIILLDFTNKDAVDWMTSOREYLLTEVGIDGFKTDGGMV 570
QY 226 WSHTHPIRHEKGVYXXXXXXXKXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVK 285
Db 571 WGRDFTFSNGEK-----GQEMR 587
QY 286 GNY-SWFVDYIYLDQMR-LNSQA-----EGKRLDVFVHWYPEAMGGGIRITNEYGNDET 338
Db 588 NRYPTDVSSYDFEAKSINPEAVSFSRSGTSCAQKSGIYV-----SGDOT 632
QY 339 K-----KARMQAPRT-----LWD-----PT-----YKEDSWIAQWNSAFILPLLR 373
Db 633 STFDSPQASLKAGLSASTSGVSYWAMDMAAGTGDYPTAEYLYKRATAMA---AFAPIMQF 688
QY 374 LKQSVDKYYPGTKLALTEYSYGGENDISGGIAMDVLGILGKNDVYMAN-----YWK 425
Db 689 HSEKSDP-----SPSEERSPWNAVARTGDETILPTFKYLYTRMNLPLYITA 736
QY 426 LKDGANNYSAYKLYRNY-DGKNA-----TFGD-----ISVNAQTS----- 461
Db 737 AKDTADNGKSMRQMAMDYPEDVNARDLDEQYMFQDGLLVAPIVQEGOTEKEVYLPESGW 796
QY 462 -DIVNSSVIASYTDASY-----KELHLIYMN-----KSMDSAFD 494
Db 797 VDIWNGVHPGGETISYADVDTLPVFAKAGAILPNMWDGQLQNGVNDLKSNDLTF 856
QY 495 AOFDLSETTYSYSGKIWGFQKNSQIKAVAPITQISGNRTYTPPLTAYHIVLTADNDT 554
Db 857 RYVP-SGDSEYS---FYDDVNGGEMRDISVSEDFANERKVSVDLPAM-----ADETT 903
QY 555 PVPPVEDPESFTLRABAGDGKVDLSWDASSGVYGVQVATYENGPFPAVASNLVETSYT 614
Db 904 MQVFESTPTSVTIDG-ADVAKADTLDAFNEATTGY-----YY 939
QY 615 DTNVTNGTSYKKITAKTKAGTSNSVLKAVPRTPVDGPDYEAEDGTGKGTIVESSGGT 674
Db 940 DT-VQNLN---YKAAAK---DAKQAILVNGVNHAP-----YAEFGHLTNTVTTASDHAG 987
QY 675 FSGTGYVTNF 684
Db 988 YTGTFGVAGF 997

RESULT 11

I39917

hypothetical protein - Bacillus sp.

C:Species: Bacillus sp.

C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 15-Oct-1999

C:Accession: I39917

R:Lee, S.

A:Title: Cloning of the aspt gene and characterization of its product, alpha-amylase-p

A:Reference number: I39917; MUID:95077378


```

Best Local Similarity 23.9%; Pred. No. 0.073;
Matches 83; Conservative 52; Mismatches 125; Indels 87; Gaps 19;

QY 400 ISGSIAMTD-----VLGILGKNDV-YMAN--YWKLKD-----GANNYV 434
      |||||      |||      |||      |||      |||      |||
Db 693 ITGGITTLDNAAKTINLKFESPYGIIYVNPDIASPAAVQWFDLIGDNTTEYGAWQYA 752
      |||||      |||      |||      |||      |||      |||
QY 435 SAAYKLYRNVGKNATFGD- ISVNAQTSDIVNSVSHASVTDSAYKELHLIVMKNKSMDSAF 493
      |||      |||      |||      |||      |||      |||
Db 753 ANALKI-----ETYGKRLVTEGTRPNITKIAAGASIGATS-----NFTAPCAY 795
      |||||      |||      |||      |||      |||      |||
QY 494 DAQFDL-----SGETTYSSGKIMGFDKNSS-----QIKAVAPITQISGNRFTYT 537
      |||||      |||      |||      |||      |||      |||
Db 796 PNQLDLRTASYTAWDNGTAV-----GFYSRRGRTCYGWFKVKTANGDGYTISEFAYN 850
      |||||      |||      |||      |||      |||      |||
QY 538 VPPLTAVHIIVLTADNDTPVPVPEDESFTLLRAEAGDGKVDLSW-DASSGVVGYSVQRATY 596
      |||||      |||      |||      |||      |||      |||
Db 851 TQPNQA--LIAGATGGITVAA---PTNLVATANSTSLQAVLTWVDNATNETGYTVERAS- 904
      |||||      |||      |||      |||      |||      |||
QY 597 ENGPFRAVA---SNLVETSYTDINVTNGTYSYYKIATKTACAGTSE-SNVLKAVPRT---- 648
      |||||      |||      |||      |||      |||      |||
Db 905 ADGYTISATLGNV--TYNTGLTAGNTYTYRVKANANAVSSDYSNVATVIEAASAY 962
      |||||      |||      |||      |||      |||      |||
QY 649 -PVDGPDRYAEADGTLKGITVESSGTSGTSGYGVTVTFNHNAGDSLTWT 694
      |||||      |||      |||      |||      |||      |||
Db 963 CTAGGSNSYIEIKDVTIGSTTTSGADASGYNYS-----KNITLT 1004
      |||||      |||      |||      |||      |||      |||

RESULT 13
S49541
cellulase - Cellulomonas fimi
C:Species: Cellulomonas fimi
C:Date: 29-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 22-Oct-1999
C:Accession: S49541; A47093
R:Meinkes, A.; Glikes, N.R.; Kwan, E.; Kilburn, D.G.; Warren, R.A.J.; Miller Jr.,
Mol. Microbiol. 12, 413-422, 1994
A:Title: Cellobiohydrolase A (Cbha) from the cellulolytic bacterium Cellulomonas
A:Reference number: S49541; MUID:94344030

```

A:Residues: 1-872 <ME2>
A:Cross-references: EMBL:L25809; NID:g456028; PIDN:AAC36898.1; PID:g456029
R.Meinke, A.; Glikes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.
J. Bacteriol. 175, 1910-1918, 1993
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
A:Reference number: A47093; MUID:93209933
A:Accession: A47093
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 812-872 <ME2>
A:Experimental source: ATCC 484
A>Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIPI:128121)
C:Superfamily: bacterial cellulose-binding domain homology
F:769-870/Domain: bacterial cellulose-binding domain homology <BCB>
E:770-869/Disulfide bonds: #status predicted

Query Match 4.1%; Score 143; DB 2; Length 872;
Best Local Similarity 21.6%; Pred. No. 0.074;
Matches 99; Conservative 42; Mismatches 151; Indels 166; Gaps 20;

QY 292 VDYLLDMLNSQAEGKRLLDVFDVH-----W-----YPEAMCGGIRITN----- 331
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 355 VNAYVDANRVDR-----VHRGAWCNPLGAGIGRFPEATSGVAASHLDAFWMI 403

QY 332 -----EVGNDETKKARMQAPRTILDWTPTYEDSKSWIAQWNSAFLLPLLPRLKQSVDK 380
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 404 KPPGESDGASTDIPNDGKRFE-----DRMCDTFE-----VSPLKNQLTG 443

QY 381 YPGTKLA-----LTEYSY---GGENDISGGIAMTDVLGIILGKNDDVTMYANKLUK 427
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 444 ATPNAPLAGQWFEEQFVTLIKVNAYPIVGTTTVEDIVAPTPTGTAGTTTATTSVPSLSWT 503

Db 444 ATPNAPLAGQWFFEEQFVTLVKKNAPYVIGTTTPVEDLVAPTVPGLTACTTATSVPLSWT 503

Tue Jul 16 15:20:07 2002

Page 9

us-09-784-554b-6.rpr

Db 1044 GSTVTAT 1050

Search completed: July 13, 2002, 00:49:03
Job time: 20159 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:59:29 ; Search time 32.61 seconds
(without alignments)
825.210 Million cell updates/sec

Title: US-09-784-554b-6

Perfect score: 3509

Sequence: 1 MKAKNSSNLSKRSKWLPPV.....SGTGYVTFNHNAGDSLMTI 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1617.5	46.1	700	1 GUNA_PAELA	P29719 paenibacill
2	1278.5	36.4	1331	1 MANB_CALSA	P22533 caldocellum
3	257	7.3	1279	1 APU_THESA	P36905 t amylpull
4	243	6.9	1861	1 APU_THETU	P38536 t amylpull
5	232.5	6.6	1475	1 APU_THETY	P16950 t amylpull
6	223	6.4	1481	1 APU_THETT	P38939 t amylpull
7	143	4.1	872	1 GUXA_CELFI	P50401 cellulomona
8	140.5	4.0	1090	1 GUXB_CELFI	P50899 cellulomona
9	136.5	3.9	828	1 PMFC_PROMI	P35114 proteus mir
10	135.5	3.9	699	1 CHIL_BACCI	P20533 bacillus ci
11	135	3.8	976	1 AMY_BUTFI	P30269 butyrivibri
12	134.5	3.8	1045	1 GUNB_CELFI	P26225 cellulomona
13	133.5	3.8	964	1 CFA2_BACCI	P70873 bacillus ci
14	129.5	3.7	3067	1 CAIC_MOUSE	Q60847 mus musculu
15	128.5	3.7	628	1 ABFA_ASPNG	P42254 aspergillus
16	127	3.6	524	1 CHID_BACCI	P27050 bacillus ci
17	124.5	3.5	747	1 GUND_CELFI	P50400 cellulomona
18	121.5	3.5	972	1 CTAL_BACCI	P94286 bacillus ci
19	119.5	3.4	1282	1 DONE_DROME	Q9vwe0 drosophila
20	118.5	3.4	481	1 PRPB_ERWCH	P16316 erwinia chr
21	118	3.4	1997	1 PRPB_HUMAN	P23467 homo sapien
22	117.5	3.3	571	1 MANB_PIRSP	P55297 piromyces s
23	116	3.3	593	1 CSG_METFE	P27373 methanother
24	115	3.3	1746	1 TENA_PIG	Q29116 sus scrofa
25	114	3.2	593	1 CSG_METSC	P27374 methanother
26	113.5	3.2	580	1 FLA2_PYRKO	Q9v2x0 pyrococcus
27	113.5	3.2	1450	1 MPFS_CHICK	Q02173 gallus gall
28	112	3.2	536	1 FLIF_BACSU	P23447 bacillus su
29	111	3.2	606	1 MANA_PIRSP	P55296 piromyces s
30	110.5	3.1	1116	1 VP2_RDVF	Q98632 rice dwarf
31	109.5	3.1	550	1 FLIC_SHIFL	Q08860 shigella fl
32	109.5	3.1	751	1 GCP2_CAEEL	P91406 caenorhabdi
33	109.5	3.1	1040	1 AXOI_RAT	P22063 rattus norv

34 109.5 3.1 1419 1 ALAL_CANAL
35 109 3.1 1433 1 SUBE_BACSU
36 108.5 3.1 1616 1 SLAP_BACCI
37 108 3.1 995 1 YIQ9_YEAST
38 108 3.1 1666 1 MYM1_MOUSE
39 108 3.1 1912 1 PTPD_HUMAN
40 107.5 3.1 524 1 GAS3_YEAST
41 107.5 3.1 619 1 CHIT_STRLI
42 107.5 3.1 1036 1 AXOI_CHICK
43 107 3.0 862 1 SLA2_BACAN
44 106.5 3.0 597 1 CHIX_STROI
45 106.5 3.0 714 1 PBPF_BACSU

ALIGNMENTS

RESULT 1
GUNA_PAELA
ID GUNA_PAELA STANDARD; PRT; 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA.
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase";
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
CC PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M76588; AAA22303.1; -.
CC PIR; B41897; B41897.
CC PIR; S27499; S27499.
CC HSP; Q06851; INBC.
CC InterPro; IPR001956; CBD_3.
CC Pfam; PF00942; CBD_3; 1.
CC ProDom; PD001947; CBD_3; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 33 POTENTIAL.
CC CHAIN 34 700 ENDOGLUCANASE A.
CC DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).
CC ACT_SITE 213 213 BY SIMILARITY.
CC SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EE0F CRC64;

Query Match 46.1%; Score 1617.5; DB 1; Length 700;
Best Local Similarity 48.7%; Pred. No. 7.4e-103;
Matches 329; Conservative 83; Mismatches 200; Indels 63; Gaps 9;

Qy 11 SKRSKWLPPVWACTIIVGGALPAPVTVHGTAKVTITKVDTSKDRKPISPIYVGTNODLA 70
 Db 3 TRQRKLFVSAALASVLTMTVPMPASVNA--AASDVTFTINTQSERAAISPIYVGTNODLS 61
 Qy 71 GDENLARLLGGNRTWYNNENNNSNAGSDWQOSSDNFLCNGGLTKAECEKPGAVTTSF 130
 Db 62 GTENWSSRLGGNRLTGYNNENNASSAGRWLHYSDDFLCNGGVPDTCDFKPGAVVTA 121
 Qy 131 HDOSLKLGAYSLVTPMAGYVAKDNGSVQESQAPARWQVYNNAKNAPFQLOPDLNDN 190
 Db 122 HDKSLGAYSVTLQWAGYVSRDKNGPVDSETPAPRWDKVEFAKNAFSLQPHLNDG 181
 Qy 191 QVYADEFVNLVKKYGAASAKAGVKGYALDNEPALMSHTHPRIHGEKVVXXXXXXXXXX 250
 Db 182 QVYMDEEVNVLNRYGNASTGTGKAYSLDNEPALMSHTHPRIHGEKVVXXXXXXXXXX 241
 Qy 251 XXXXXXXXXXXXXXXXXXXXXXXXFKLXI--GTSVKGNYSWFVYDYLDMRLNSQAEK 308
 Db 242 SKAVKNVDPHAEIFGPAHYGFGAYLSQDAPGWPSPGNGYSWFIDYLDQMKNAHTQNGK 301
 Qy 309 RLLDVFVHWYPEAMGGIRIT--NEVGNDETKKARMOAPRTLMDPTYKEDSWIAQWNSAF 367
 Db 302 RLLDVLVHWYPEAQGGQRIYVGGAGNIDTKARQVAPRSLDPAQOEDSWIGTWFFSY 361
 Qy 368 LPLPLRLKQSDVKYPTGKALTEYSYGGENDISGGTAMTDVILGKNDVYVANYWKLK 427
 Db 362 LPLPKLQSSQTQYPTGKALTEYSYGGENDISGGTAMTDVILGKNDVYVANYWQTE 421
 Qy 428 DGANNYSAAKLYRNDGKNATGDI SVNAQTSDI VNSSVHASVTDASVYKELHLVMMK 487
 Db 422 DN--TDYTSAAKLYRNDGKNATGDI SVNAQTSDI VNSSVHASVTDASVYKELHLVMMK 480
 Qy 488 SMDSAFAQDFLSETTYSYSGKIMGFDKNSQIKAVAPITQISGNRTYVPPPLTAYHIV 547
 Db 481 NFDDPINATFQSLGDKYTSRQWGFDTGSDITEQAATINNNQTYTLPLSAHVIV 540
 Qy 548 LTANDTPV-----PVEDPESETLR----- 568
 Db 541 LKADSTEPVNDLVQYKDGDRNATDNQIKPHFNIGNKGTSPV-DLSSLTRYETKDS 599
 Qy 569 -----AEAGDGKVLSDWASSGV-----VGYSVQRATY-ENGPFPAVASNLVE 610
 Db 600 SAANNGWIDWAKLGSNIQISFGNHNGADSDTYAELGFSGAGSIABGGSGEIQLRMSK 659
 Qy 611 TSYTDNTVNTGTSY 624
 Db 660 ADWSNEANDYSF 673
 RESULT 2
 MANB_CALSA STANDARD; PRT: 1331 AA.
 AC P22533;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
 beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
 mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
 GN MANA.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/clostridium group;
 OC Thermoanaerobacter group; Caldicellulosiruptor.
 OX NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93119139; PubMed=1476429;
 RA Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
 multi-domain enzyme."
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).
 RN [2]
 RP SEQUENCE OF 1-346 FROM N.A.

RX MEDLINE=91247819; PubMed=2039230;
 RA Luthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 gene coding for a beta-mannanase from the extremely thermophilic
 bacterium 'Caldocellum saccharolyticum'";
 RT Appl. Environ. Microbiol. 57:694-700(1991).
 RL [1]
 CC -!- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
 POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH
 MANNAASE AND ENDOGLUCANASE ACTIVITIES.
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 linkages in mannans, galactomannans, glucomannans, and
 galactoglucomannans.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -!- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
 CELSIUS.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L01257; AAA71887.1; -
 DR EMBL: M36063; AAA72861.1; -
 DR PIR: B43745; B43745.
 DR PIR: A48954; A48954.
 DR HSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001547; Glyco_hydro_F5.
 DR Pfam: PF00942; CBD_3; 2.
 DR Pfam: PF00150; cellulase; 1.
 DR ProDom: PD001947; CBD_3; 2.
 DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 KW Hydrolase; Glycosidase; Cellulose degradation; Signal;
 KW Multifunctional enzyme.
 KW SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 FT SEQUENCE 1331 AA; 146892 MW; FFBCA51BB8DF0E0 CRC64;
 Query Match 36.4%; Score 1278.5; DB 1; Length 1331;
 Best Local Similarity 45.7%; Pred. No. 2.3e-79;
 Matches 249; Conservative 85; Mismatches 190; Indels 21; Gaps 6;
 QY 32 PAPTVMHGTAKVTITKVDTSKDRKPISPIYVGTNODLAGDENLAARRLGGNRTWYNN 91
 Db 777 PLPTI--SPSPSVETINTNAGRTQISPIYIGANQIEGVVH--SARRLGNRLTGYNWE 833
 QY 92 NNSNAGSDWQOSSDNFLCNGGLTKAECEKPGAVTTSFHDOSLKLGAYSLVTPMAGV 151
 Db 834 NFNNSNAGNDWYHSSDDYLCWSMGISGEDAKVPAVAAVSKFHEYSKLNAYSAVTLQMGV 893
 QY 152 AKDNGSVQESQAPARWQVYNNAKNAPFQLOPDLNDNQVYADEFVNLVKKYGAASAK 211
 Db 894 SKDNYGTVSENETAPSNRWAEEKFKKDAPLSLNPDNDNFVYDDEFINLYNKYGNASSP 953


```

RESULT 4
APU_THETU
ID APU_THETU STANDARD; PRT; 1861 AA.
AC P38536;
DT 01-OCT-1994 (Rel. 30, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase
DE type II) [Includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan
DE glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan
DE glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].
GN AMB.
OS Thermoanaerobacter thermosulfurogenes (Clostridium
OS thermosulfurogenes).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacterium.
OX NCBI_TaxID=33950;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 3896 / EMI;
RX MDLINE=94252998; PubMed=8195085;
RA Matschek M., Burchardt G., Sahm K., Bahl H.;
RT "Pullulanase of Thermoanaerobacterium thermosulfurogenes EMI
RT (Clostridium thermosulfurogenes): molecular analysis of the gene,
RT composite structure of the enzyme, and a common model for its
RT attachment to the cell surface."
RL J. Bacteriol. 176:3295-3302(1994).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha and beta-limit dextrins of amylopectin and glycogen.
CC -I- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN
CC S-LAYER ANCHOR.
CC -I- PTM: GLYCOSYLATED.
CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOG (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57692; AAB00841.1; -
DR HSSP; Q08751; 1BV2.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001119; SLH.
DR InterPro; IPR004185; alpha-amylase_N.
DR InterPro; IPR004193; isoamylase_N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase_C; 1.
DR Pfam; PF02903; alpha-amylase_N; 1.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF02922; isoamylase_N; 1.
DR Pfam; PF00395; SLH; 3.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 3.
DR Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
KW Multifunctional enzyme; Glycoprotein.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1861 AMYLOPULLULANASE.
FT DOMAIN 928 1018 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1157 1248 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 628 628 BY SIMILARITY.
FT ACT_SITE 657 657 BY SIMILARITY.
FT ACT_SITE 734 734 BY SIMILARITY.
FT DOMAIN 1681 1739 SLH 1.

```

```

FT DOMAIN 1740 1803 SLH 2.
FT DOMAIN 1804 1861 SLH 3.
FT CONFLICT 1734 1734 D -> E (IN REF. 1: AAB00841).
SQ SEQUENCE 1861 AA; 206104 MW; 06C23070EA53B574 CRC64;

Query Match 6.9%; Score 243; DB 1; Length 1861;
Best Local Similarity 20.8%; Pred. No. 1.5e-08;
Matches 138; Conservative 86; Mismatches 228; Indels 210; Gaps 30;

QY 62 IYGTNQDLADENLAARLGGNRMTGYWNNMSNA-----GSDWQSSDNFLCN 111
DB 497 MFGTTQD-----FEKLMSDAHAKGIKIILDGVFNHTSDSYFN 535
QY 112 NGLLTKAECCKPGAVTTSFHDQSLKLGAYSLVTLPMAGYVAKDNGSVQSEQAARWN 171
DB 536 RYG-----KYPG-----LGAYQ-----AWKEGNSL-----SPYGDW- 562
QY 172 QVYNAKNA-----PFQLQPLDNDNOVYADEFVNFVLVKKYGAAS-----TKAGV 214
DB 563 YTINSDGTTECWGVDYSLPVIKSLNGSEYNVTSWANFIINDENAIKYWLNPDGNLNDGA 622
QY 215 KGYALD--NEPA--LWSHTHPRIHGEKVKYXXXXXXXKXXXXXXXXXXXXXXX 270
DB 623 DQWRLDVENVAHDFWTHFRNAINTVKFEAPMAIENWGDA----- 662
QY 271 XXXXFKLHXIGTSVKRG--NYSW--FVDYVLDQM-----RLNSOAEGRLLDVFV 316
DB 663 -----SLDLGDSFNSVMNYQFRNDIIDLIGQSDGNGQHNPIDAAKLQRLMSIYER 717
QY 317 HWYPAMGGGIRITNEVGNDETTKARMOQAPRTLWDPTYKEDSWIAQMNSAFPLPLPKQ 376
DB 718 YPLPAFYS---TMNLGSHDTMRLITVFGYNSADPNENSDA-AKQLAEQKLKLTILQM 772
QY 377 SVDKYYPGKTKALTEYSYGGENDISGGIAMDVLGI-LGKNVYMANVYK----- 425
DB 773 G-----YPG-----MADYYGDEAGVSGGKDPDRRTFPMGNETTLQDFKNISSIRNNQ 824
QY 426 -LKDGANNVVSAAVKLY----RNYDGKNATFGD-----ISVNAOTSD--IVNSSVHA 470
DB 825 VLKTDLETLYAONDVYAGRIIRLNGKDA-FGTSYPSDAAIVAINRSKDKQIAIDITKF 883
QY 471 SVTDASYKELHLIVYMNKSDSAFDAQFDLSGETTYSYSSKINGFKNKSQIKAVAPITQIS 530
DB 884 LRDGVTFFDL-----INNVSYS-----IS 903
QY 531 GNRTYTYVPLTAYHIVLTADNDTPVPVPEDEPSEFTLRAEAGDKVDLSWDASSGVVGVYS 590
DB 904 NGQIVIDVPAMSG--VMLISDDGQDLTAPQAPSNVVV--TSGNGKVDLSWLQSDGATGYN 959
QY 591 VORATYENGPFPAVASNLVETSYDTNTVNGTSYKYKITAKTGSTP--SNVLKAVPRT 648
DB 960 IYRVSVEGGLYEKIASNVTTETTFEDANVTNGLKYVYAIASIDELGNESGISNDAVAVPAY 1019
QY 649 PV 650
DB 1020 PI 1021

RESULT 5
APU_THETU
ID APU_THETU STANDARD; PRT; 1475 AA.
AC P16950;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Amylopullulanase precursor (Alpha-amylase/pullulanase) [Includes:
DE Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);
DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)
DE (Alpha-dextrin endo-1,6-alpha-glucosidase)].
GN APU.
OS Thermoanaerobacter thermohydrosulfuricus (Clostridium
OS thermohydrosulfuricum).

```


OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoanaerobacter group; Thermoanaerobacter.
 OX NCBI_TaxID=1516;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E101-69;
 RX MEDLINE=90362027; PubMed=2391488;
 RA Melasniemi H., Palohimo M., Hemioe L.;
 RL "Nucleotide sequence of the alpha-amylase-pullulanase gene from
 RT Clostridium thermohydrosulfuricum";
 RL J. Gen. Microbiol. 136:447-454 (1990).
 RN [2]
 RP SEQUENCE OF 32-39.
 RC STRAIN=E101-69;
 RX MEDLINE=88268757; PubMed=3260488;
 RA Melasniemi H.;
 RL "Purification and some properties of the extracellular alpha-amylase-
 RT pullulanase produced by Clostridium thermohydrosulfuricum";
 RL Biochem. J. 250:813-818 (1988).
 RN [3]
 RP FIBRONECTIN TYPE III DOMAINS.
 RX MEDLINE=93028390; PubMed=1409594;
 RA Bork P., Doolittle R.F.;
 RL "Proposed acquisition of an animal protein domain by bacteria";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:8990-8994 (1992).
 CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
 CC linkages in pullulan and in amylopectin and glycogen, and the
 CC alpha- and beta-limit dextrins of amylopectin and glycogen.
 CC -|- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -|- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M28471; AAA23205.1; -
 DR PIR; A44765; A44765.
 DR HSSP; 008751; 1BVZ.
 DR InterPro; IPR000461; Alpha_amylase.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR004185; alpha-amylase_N.
 DR Pfam; PF00128; alpha-amylase; 1.
 DR Pfam; PF02806; alpha-amylase_C; 1.
 DR Pfam; PF02903; alpha-amylase_N; 1.
 DR Pfam; PF00041; fn3; 2.
 DR SMART; SM00060; FN3; 2.
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
 KW Multifunctional enzyme.
 FT SIGNAL 1 31
 FT CHAIN 32 1475 AMYLOPULLULANASE.
 FT DOMAIN 927 1016 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 1163 1255 FIBRONECTIN TYPE-III 2.
 FT ACT_SITE 629 629 BY SIMILARITY.
 FT ACT_SITE 658 658 BY SIMILARITY.
 FT ACT_SITE 735 735 BY SIMILARITY.
 SQ SEQUENCE 1475 AA; 165631 MW; 3476C414110BE376 CRC64;
 Query Match 6.6%; Score 232.5; DB 1; Length 1475;
 Best Local Similarity 22.8%; Pred. No. 5.6e-08;
 Matches 140; Conservative 69; Mismatches 190; Indels 215; Gaps 33;
 QY 132 DOSLKLGYAS--LVTLPMAGYVA-RDNGSVQSEQAAPSARNQVNVNAKNAFFQLQPD-- 186
 DB 527 DOSIYFDRYKGLNTGVLGAYQAWKQDQS-----KSPYGDW-----YEIKPDGT 571

QY 187 -----LNDNQVYADEFNFLVKYGAAS-----TKAGVKGYALD-- 220
 DB 572 YEGWGFDSLPVIRIQINGSEYNVKSADFIINNPNALISKYWLNDGDKNVGADGWRDLVA 631
 QY 221 NEPA--LMSHTPHRIHGEKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXSLD 666
 DB 632 NEVAHDFWVHFRGAINTVKPNAPVAENWDA-----SLD 666
 QY 279 XIGTSVKG--NYSW---FVDYLLQD-----RLNSQAEGRKLLDVFVHVHP----- 320
 DB 667 LLGDSFNSVMNYLFRNAVVIDFILDKSFDDGNGVNVHNPIDAAKLDQRLMSIYERYPLPVFYS 726
 QY 321 --EAMGG--GIRITNEVG--NDETKKARMAQAPRTLWDPTYREDSWIAQNSAFLPULPLK 375
 DB 727 TMLLGSHTDTRILTVFGYNSADENQNSQAADL-----AVRLK 766
 QY 376 QS--VDKYYPGCKLALTEYSYGGENDISGG-----IAMDVLGILG 414
 DB 767 LAAILQMGYPC---MPSIYYGDEAGSGGKDPNRRTPFWGREDTDLQTFKKVNVNIRN 822
 QY 415 KNDVYMANWKLDGANNVSAAYKLYRNDYDKNATFGDISVNAQTSIDIVNSSVHASVTD 474
 DB 823 ENQV-----LKTG-----DLETLYAN--GDVYAFGRRIINGK--DTEGKSYDPSVA- 864
 QY 475 ASYRELHLIVMVK-----SMDS-----AF-DAOFDLSGET-TYSSGKINGFDKNSS 518
 DB 865 -----IVINKGDAKQVSDITTKFIRDGVAFTDA--LSGKTYTVQDGKI----- 906
 QY 519 QIKAVAPITQISGRNFTYTPPLTAYHIVLTADNDTPVPVPEDEPSETLRAEAGDKVDL 578
 DB 907 ---VVEGSGMDG-----AILSDTGQNLTAQPIID-----LRAVSGNGKVDL 946
 QY 579 SWDASSGVGVSVORATYENGPPFAAVALNSLVETSYDTNTNGTSYTYKKTAKTACTSE 638
 DB 947 SWSVVDKAVSYNRYSTKGLYEKIASNVITQITDTTEVTNGLKYVAYTAVDNDGNS 1006
 QY 639 --SNVLKAVRTPV 650
 DB 1007 ALSNEVEYAPFPI 1020
 RESULT 6
 APU_THEET
 ID APU THEET STANDARD; PRT; 1481 AA.
 AC P38939;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Amylopullulanase precursor (Alpha-amylase/pullulanase) [Includes:
 DE Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase);
 DE Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan glucanohydrolase)
 DE (Alpha-dextrin endo-1,6-alpha-glucosidase)].
 GN APU.
 OS Thermoanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoanaerobacter group; Thermoanaerobacter.
 OX NCBI_TaxID=1757;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=39E;
 RX MEDLINE=90147689; PubMed=2302196;
 RA Mathupala S.P., Saba B.C., Zeikus J.G.;
 RL "Substrate competition and specificity at the active site of
 RT amylopullulanase from Clostridium thermohydrosulfuricum";
 RL Biochem. Biophys. Res. Commun. 166:126-132 (1990).
 CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -|- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
 CC linkages in pullulan and in amylopectin and glycogen, and the
 CC alpha- and beta-limit dextrins of amylopectin and glycogen.
 CC -|- SUBUNIT: MONOMER.
 CC -|- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.

```
CC CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; M97665; AAA23201.1; -.
DR DR PIR; S28669; S28669.
DR DR HSSP; Q08751; 1BVZ.
DR DR InterPro; IPR000461; Alpha_amylase.
DR DR InterPro; IPR003961; FN_III.
DR DR InterPro; IPR004185; alpha-amylase_N.
DR DR Pfam; PF00128; alpha-amylase; 1.
DR DR Pfam; PF02806; alpha-amylase_C; 1.
DR DR Pfam; PF02903; alpha-amylase_N; 1.
DR DR Pfam; PF00041; fn3; 2.
DR DR SMART; SM00060; FN3; 2.
KW Hydrolyase; Glycosidase; Carbohydrate metabolism; Signal; Repeat;
KW Multifunctional enzyme.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 1481 AMYLOPULLULANASE.
FT DOMAIN 926 1015 FIBRONECTIN TYPE-III 1.
FT DOMAIN 1158 1250 FIBRONECTIN TYPE-III 2.
FT ACT_SITE 628 628 BY SIMILARITY.
FT ACT_SITE 657 657 BY SIMILARITY.
FT ACT_SITE 734 734 BY SIMILARITY.
FT ACT_SITE 1481 AA; 166362 MW; CB2635960539CAID CRC64;
SQ SEQUENCE 1481 AA; 166362 MW; CB2635960539CAID CRC64;

Query Match 6.4%; Score 223; DB 1; Length 1481;
Best Local Similarity 21.3%; Pred. No. 2.5e-07;
Matches 125; Conservative 77; Mismatches 186; Indels 200; Gaps 28;

QY 154 DNGSVQESQAPSRNQVNAKNAP---FQLQPD-----LNDNQVY 193
DB 533 DRYGKYLDNELGAYQAWKQDQSKSPGYDWEYKIPDGTVEGMWGFSLPVIQINGSEYN 592
QY 194 ADEPVELVKYGAAS-----KAGVKGYALD--NEPA--LWSHTHPRIHGEKXX 240
DB 593 VKSWADFIINPNNAISYKWLNPDCRDAGADGRLDVANEIAHDFVHFAAINTVKPNA 652
QY 241 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXKHLXIGTSVKG--NYSW---FVDYY 295
DB 653 PMIAELMGDA-----SLDLLGDSFNSVMYLFPRNAVIDFI 687
QY 296 LDQM-----RLNSQAEGRLLDVFVHWP-----EAMGG-----GIRIT 330
DB 688 LDKQFDGQNVVHPIDAAKLDQRLMSIERYPLVPFYSTMNLGSHDTMRILTVFGYNSA 747
QY 331 NEVGNDETAKRQAAPTLDPTYKEDSWIAQNSAFLLPLRLKQSVKYPYPGTKLALT 390
DB 748 NE--NQNSQAKDLAVKRL-----KLAAILQMGTPG-----MP 778
QY 391 EYSGGENDISGGIAMTDVLGI--LGKNDVYMYNYK-----LKDGNANNYSAA 438
DB 779 SIYVGDEAGQSGKDPNRRFTFSWGREDDQLQDFKVKVYNNRNEQVLTKG-----DLE 832
QY 439 KLYRNDGKNATGDISVNAQTSDIYNSVSHASVTDAKYKELHIVNKSMSAFDAQPD 498
DB 833 TLXAN--GDVYAFGRRLINGK--DFVGNYPDSVA-----IVVINKG--EAKSVQID 878
QY 499 -----LSGET--TYSGGKIGWFDKNSSQIKAVAPITQISGNRFTYTPPLTA 543
DB 879 TTFVRCGVAFTDALSGKTYTVDGQI-----VVEVALDG-----914
QY 544 YHIVLTAD---NDTPVPVPEDESFTLRABAGDKGVLDLSDWASSGVVGYSVQATYENGP 600
DB 915 --AIIISDPQNLTAPOITD-----LKAVSGNGQVDLSKSAVDRAVSYNIYRSTVKGL 967
```

```
QY 601 FAAVASNLVETSYTDNTVNGTSYVYKTKACTSE--SNVLKAVP 646
DB 968 YEKIASNVQTQIYTDVTNGLKYVSVTAVDSDGNSALSNEVEAYP 1015

RESULT 7
GUXA_CELFI STANDARD; PRT; 872 AA.
ID GUXA_CELFI
AC P50401;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Exoglucanase A precursor (EC 3.2.1.91) (Exocellobiohydrolase A)
DE (1,4-beta-cellobiohydrolase A) (CBP95).
GN CBHA.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 484;
RX MEDLINE=94344030; PubMed=8065260;
RA Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA Miller R.C. Jr.;
RT "Cellulobiohydrolase A (Cbha) from the cellulolytic bacterium
RT Cellulomonas fimi is a beta-1,4-exocellobiohydrolase analogous to
RT Trichoderma reesei CBH II."
RL Mol. Microbiol. 12:413-422(1994).
RN [2]
RP SEQUENCE OF 41-58.
RX MEDLINE=93209933; PubMed=8458833;
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RT D (Cenb), a family A beta-1,4-glucanase."
RL J. Bacteriol. 175:1910-1918(1993).
CC -!- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF
CC CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCELLULOSE, BACTERIAL
CC ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION
CC OF ANOMERIC CONFIGURATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; L25809; AAC36898.1; -.
DR DR HSSP; P07986; 1EXG.
DR DR InterPro; IPR001919; CBD_2.
DR DR InterPro; IPR003961; FN_III.
DR DR InterPro; IPR003962; FNIII_repeat.
DR DR InterPro; IPR001524; Glyco_hydro_6.
DR DR Pfam; PF00553; CBD_2; 1.
DR DR Pfam; PF00041; fn3; 3.
DR DR Pfam; PF01341; Glyco_hydro_6; 1.
DR DR PRINTS; PR00014; FNTYPEIII.
DR DR PRODOM; PD00733; GLHYDRLASE6.
DR DR PRODOM; PD003733; Glyco_hydro_6; 2.
DR DR SMART; SM00060; FN3; 3.
DR DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
```


QY 433 YVSAAYKLYRNDGKNATPGDISVNAQTSDIVNSSVHASVTDASYKELHLIVNKNKMSDSA 492
 Db 766 YTVRAFAAGNVPASAL-TVTTKATPSTTAPSPAITSSSTVNTSIVTIGWSASTDNA 824
 QY 493 FDAQDLSGGETTSSGKINGFDKNSQIKAVAPIT-----QISGNFTYTV----- 538
 Db 825 GGS--GLAGDVI-----RGATRAVQTTALTFTDGLTASTAYEYTVRARDVAGNV 873
 QY 539 -PPLTAYHIVLTADN--DTPVPPVEDPESFTLRAEAGDKVLDSDAS-----SGVVGYS 590
 Db 874 SAPSTAVSVTTKSDTTPDTAPSV--PAGLAAMT-VTETSVALTWNASTDTGGSLKGID 930
 QY 591 VQRTYENGEPFAAVASNLVETSYDTNVTNGTSYKYITAKTRAG--TSESNNVLKAVPRT 648
 Db 931 VYRG-----ATRVGSTTTASVTDGLTAATAYQYTVRATDNAGNVSAASAALSVTTKT 983
 QY 649 PVDGPD--RYEA--EDGILKGTIVESSCT-----GFSGT-----GYVTNFHNA 687
 Db 984 PQTGGSCSVAYNASSWNSGFTASVRINTGTTINGWSLGLDFTAGOKVQOGWSATWTQS 1043
 QY 688 GDSLTMT 694
 Db 1044 GSTVTAT 1050

RESULT 9
 PMFC_PROMI STANDARD; PRT; 828 AA.
 AC P53514;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Outer membrane usher protein pmfc precursor.
 GN PMFC.
 OS Proteus mirabilis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Proteus.
 OX NCBI_TaxID=584;
 RN [1]
 RC STRAIN=HI4320;
 RC MEDLINE=95047519; PubMed=7959033;
 RA Massad G., Mobley H.L.F.;
 RT Genetic organization and complete sequence of the Proteus mirabilis
 pmf fimbrial operon.;
 RL Gene 150:101-104(1994).
 CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PMF FIMBRIAL
 SUBUNITS ACROSS THE OUTER MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z35428; CAA84590.1;
 DR InterPro: IPR000015; FimB_usher.
 DR Pfam: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 828 OUTER MEMBRANE USHER PROTEIN PMFC.
 SQ SEQUENCE 828 AA; 93108 MW; B11B05992A64BA44 CRC64;

Query Match 3.9%; Score 136.5; DB 1; Length 828;
 Best Local Similarity 19.6%; Pred. No. 0.091;
 Matches 145; Conservative 93; Mismatches 243; Indels 259; Gaps 44;

QY 34 PTVHVGQTAQTVTIKVDTSKDRKPISPYIYGTNODLAGDENLAARRLGG-----NRMT 86
 Db 97 PTL-----KLTINQENQYIKQISQPCDFISQ-LPG---ISIKNDGGVLDITLPRSLM 147
 QY 87 GYN-----WENMSNAGSDWQSSDNFLCNGGLTKACEKPGAVTTSFHDQSLKL 137
 Db 148 KYEESDTPPELWDSVGLIFDYTLTGTSTRPNKGNNN-----TLTYGQAGLNL 199
 QY 138 GAYSLVTLPMAGYVAKDGNQVSESEQAPSAR--WNQVVNAKNAPFQLQPLDNDNQVAD 195
 Db 200 GEWRL-----RAY-----QGNYS---SEYSSNRFQWQIYAYKPLP-DLAAKLTVGETYLN 248
 QY 196 E--FVNFLVKYKGAASTKA-----GVKGYALDNEPALWSTHPIHGEKVVXXXXXXX 246
 Db 249 SQIFDSF---RFTGANLQSDERMLPPSLQGYA-----PEIHG----- 282
 QY 247 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXICTSVKGNYSWEVDYVLDQMLNSOAE 306
 Db 283 --IANTNAKVTVTQNGRLIYETVPAGPEVINHLQNTVQGG-----LD-VRVEEQ-N 330
 QY 307 GKRLLDVFDVHW--YP-EAMGGGIRITNEVGNDETKKARMOAPRTLWDPTYKEDSWIAQW 363
 Db 331 GK--INEFOQTANLPYMTRPQSVRENTSLGQSSVNNHMKQGP----- 371
 QY 364 NSAFLLPLRLKQSVDKYYPGTKLALTEYSYGEN--DISGGIAMTDLVLGILKNDVYMA 421
 Db 372 -----LFYQG-----DFSNGMNTWSLYGGTLLT-----AKD---Y 399
 QY 422 NYWKLKDG-----ANNYSAAKYLRNYDGKNATFGDISVN-AQTSDIVNSSVHASV 472
 Db 400 NAWSLGIGHDMGRFGTSLGSDITQYS--KTYDNEKINGMSFKLNYAKTFD---EYHSTI 453
 QY 473 TDASY-----KELHLIVMKNKMSDSAFAQFDLSGETTY 505
 Db 454 TFAGYRFEKTPRSPFSQYIDERYNINNGYEKEMYTITGNKTF-WADDAEKSITLYLSY 512
 QY 506 SSGKIWGFDKNS-----SQIKAVAPITQISGNRFTYTPPLTAYHIVLTADNDTPVP 557
 Db 513 RHQYW--DKNTQEQYGVTVSRNFSIMGIEQINTN-----LSAFRTQYKGTND--- 558
 QY 558 PVEDPESFTLRAEAGDKVLDSDASSGVVGSVQVATYENGPPFAAVAS-----NL- 608
 Db 559 ---DPLSNISLPLSGSR-----NIGYNLQD---NNGKVTQMASYADNRDYNLW 602
 QY 609 -VETSYTDTNVTNGTSYKYITAKTKAGTSNSVLKAVPRTVPDGPDRYEADGLTKGTI 667
 Db 603 RIRAGLSDDKKANTDGY-----QHSQYAEINANASYQQ-----DNYLAVGATIKG-- 649
 QY 668 VESSGTGSGTGYVTFHNA 687
 Db 650 -----GFTATRYGAALHSS 663

RESULT 10
 CHIL_BACCI STANDARD; PRT; 699 AA.
 ID CHIL_BACCI
 AC P20533;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Chitinase A1 precursor (EC 3.2.1.14).
 GN CHIAL.
 OS Bacillus circulans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WL-12;
 RX MEDLINE=90368776; PubMed=2203782;
 RA Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;
 RT "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed

its evolutionary relationship to Serratia chitinase and to the type III homology units of fibronectin";
 J. Biol. Chem. 265:15659-15665(1990).

[2] MUTAGENESIS.
 RC STRAIN-WL-12;
 RX MEDLINE-93366760; PubMed-8103047;
 RA Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,
 RA Uchida M., Tanaka H.;
 RT Identification of glutamic acid 204 and aspartic acid 200 in
 RT chitinase A1 of Bacillus circulans WL-12 as essential residues for
 RT chitinase activity.;
 RL J. Biol. Chem. 268:18567-18572(1993).
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
 CC acetyl-D-glucosamine polymers of chitin.
 CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -I- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
 CC HYDROLASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

EMBL; M57601; AA81528.1; -
 DR PIR; A38368; A38368.
 DR HSP; P07254; ICTN.
 DR InterPro; IPR003610; Chitin_bind3.
 DR InterPro; IPR001579; Chitinase_2.
 DR InterPro; IPR003961; FN.III.
 DR InterPro; IPR003962; FN.III.repeat.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR Pfam; PF02839; CBD_7; 1.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR PRINTS; PR000114; FNTYPEIII.
 DR SMART; SM00495; ChitBD3; 1.
 DR SMART; SM00060; FN3; 2.
 DR PROSITE; PS01095; CHITINASE_18; 1.
 DR Hydrolase; Glycosidase; Chitin degradation; Signal; Repeat.
 FT SIGNAL 1 41
 FT CHAIN 42 699
 FT DOMAIN 42 460
 FT DOMAIN 465 549
 FT DOMAIN 560 644
 FT ACT_SITE 204 204
 FT MUTAGEN 200 200
 FT MUTAGEN 200 200
 FT MUTAGEN 204 204
 FT SEQUENCE 699 AA; 73677 MW; AC7C9B2E2987643 CRC64;

Query Match 3.9%; Score 135.5; DB 1; Length 699;
 Best Local Similarity 20.9%; Pred. No. 0.085;
 Matches 114; Conservative 69; Mismatches 145; Indels 217; Gaps 32;

QY 286 GNYSW-----FVYLLDQMLNSQAEGRLLDVFVHW-YPEAMGGGIR 328
 DB 162 GQWTSNRFSDVAATAATREVFANSVDFLR-----KYNFDGVDLDWEYp--VSGGLD 212
 QY 329 ITNEVGNDKTKARMQAPRTLLDPPYKEDSWIAQNSAFLLPLRLKQSDVK--YYPGTK 386
 DB 213 -----GNSKR-----PEDQN-----YTLKSKREKLDAAAGVDGKK 245
 QY 387 LALT-----EYSGGENDISGIAITDVLGIL-----G 414
 DB 246 YLLTIASGASATYAANTELAIAIVDWINITWDFNCAGWKISAHNAPLNYDPAASAAG 305
 QY 415 KNDVTMANYWKLKGCANVYSA---AYKLY-----RNYDG----- 446

Db 306 VPD---ANTFNVAAGAOGLHDAGVPAAKLVLCVPFFYGRGWDGCAQAGNGOYQTCGGSSV 362
 QY 447 -----KNATFGDISVNAQTSDIVNSSVHASVTDASYKELHLI-VMNKSMDSAFDAQFDLS 500
 Db 363 GTWEAGSDFDYDLEAN-----YINKGYTRYWNTAKVPYLYNASKNRKFISYDDAE-SVG 416
 QY 501 GETTYSCK-----IWGF-DKN---SSQIKAVAPITQISGNRFTTVVPLTAYHIVLT 549
 Db 417 YKTAIYIKSGLGGMFWELSGDRNKTLOKLIKADLP-----TGG-----TVPPV----- 460
 QY 550 ADNDTPVPVPE-DPESFTLRAEAGDKVDLSWDASS---GVVGYSVQRAIVYENGPFPAVA 605
 Db 461 ---DTPSPVPGNARSTGVANS-----VTLAWNSTDNVGVYGVNY-----NG--ANLA 506
 QY 606 SNLVETSYTDNVTNGTSYYKIKAKTAKG--TSESNVLA-----VPRTPVD- 651
 Db 507 TSVTGTATTISGLTAGTSYFTTIKAKDAAGNLSAASNAVTVSTTAQPGDGTQAPTPTNL 566
 QY 652 -----GPDYEAEDGTLKGTIVSESGTGFSGTGYVTFHNAGDS 690
 Db 567 ASTAQTTSSTLSWTASTDNVGVYGVYNGTALATTVTGTATTISGLA-----ADTS 619
 QY 691 LTMTI 695
 Db 620 YTFV 624

RESULT 11

AMY_BUTFI
 ID AMY_BUTFI STANDARD; PRT; 976 AA.
 AC P30269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-amylase precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
 DE glucanohydrolase).
 GN AMVA.
 OS Butyrivibrio fibrisolvens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Butyrivibrio.
 OX NCBI_TaxID=831;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HI7C;
 RC MEDLINE-91286207; PubMed-2061294;
 RA Rumbak E., Rawlings D.E., Lindsey G.G., Woods D.R.;
 RT "Cloning, nucleotide sequence, and enzymatic characterization of an
 RT alpha-amylase from the ruminal bacterium Butyrivibrio fibrisolvens
 RT HI7C.";
 RL J. Bacteriol. 173:4203-4211(1991).
 CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
 CC linkages in oligosaccharides and polysaccharides.
 CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M62507; AAA23005.1; -
 DR PIR; A42466; A42466.
 DR HSP; P00691; 1BAG.
 DR InterPro; IPR000461; Alpha.amylase.
 DR Pfam; PF00128; alpha-amylase_1.
 DR Pfam; PF02806; alpha-amylase_C; 1.
 DR Hydrolase; Glycosidase; Carbohydrate metabolism; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 976
 FT ACT_SITE 323 323 BY SIMILARITY.

Db 313 KYDDANRWLDYTVGVNGKVPYS-----PGQAVLDSWALRYAANTSFVALVYSDWM 367
QY 210 TKA-----GVK--GVALDNEPALWSHT-----HPRIHGEKVXXXXXXX 246
Db 368 TDATRKARYHDFGVQINVALGONPRSSSVVGVGANGPTAPHRHTAHG----- 416
QY 247 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNYSWFVYDLDQMLNSQAE 306
Db 417 -----SW-----LDSITTPAQR 429
QY 307 GKRLDQVDFVHTPEAMGGGIRITNEVGNDKTKKQAPRTLMDPTYKEDSWIAQWNSA 366
Db 430 -----HVLGALVGGPGSPNDAYTDSRQ-----DYVANEVATDYNAG 466
QY 367 FLPLLPRLKQSVKYPGKTLA--LTEXSYGGE-----NDISGIAWTDVLGILGN 416
Db 467 FTSALARLVEE---YGGTPLASFPTEPQPDGDLFVAMLNQPPSG-TFTEVKAMIRNQ 521
QY 417 DVYMA-----NYMKLDGANNYSAAKLYRNY-----DGKNATFGDISVNAQTS 462
Db 522 SAPPARSLKNAKYWFETDGG---FAASDVTLSSANYSECGAQSGKVSAGGTILGYVELS- 577
QY 463 IVNSVHVASVTDASYKELHLIV-----MKNKSMDSAF-----DAQFDLSGETTY-SSGKTW 511
Db 578 CVGQDIHPGQSQHRREIOFRLTGPAGWNPANDPSPYGLTQTALAKASAITLYDGSGLVM 637
QY 512 GFDKNSQIKAVAPITQISGNRTYTPPLTAVHIVLTADNDTPVPVDPEDPESFILRAEA 571
Db 638 GKPEGTGT-----TDTTPP-----TTPGTPPATVGT-TVGA-- 667
QY 572 GDGKVDLSW---DASSGVVGVSVQVATYENGPFPAASVNLVETSYTDTNVTNGTSYK 627
Db 668 -----SLSWAASDAGSVAGYELRV---QGTQTLVGTTTAAVILDLTPGAYSV 719
QY 628 ITAKTRAGTSESNVLKAVRTPVDGPDYAEADGTLKGITVSESSGFGSGT 678
Db 720 VKARDVAGNVSAAASAAVTTDTTG-----ETEPPTPGTPVASAVTSTGAT 766

RESULT 13
CTA2_BACCI
ID CTA2_BACCI STANDARD; PRT; 964 AA.
AC P70873;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cycloisomaltooligosaccharide glucanotransferase precursor (EC 2.4.1.-)
DE (CITASE).
GN CIT.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U-155;
RA Oguma T., Kurokawa T., Tobe K., Kitao S., Kobayashi M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D88360; BAAL3595.1; -
KW transferase; Glycosyltransferase; Signal.
FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 964 CYCLOISOMALTOOLIGOSACCHARIDE
FT GLUCANOTRANSFERASE.
SQ SEQUENCE 964 AA; 107208 MW; 8849CD5E2DE9A68 CRC64;

Query Match 3.8%; Score 133.5; DB 1; Length 964;
Best Local Similarity 22.5%; Pred. No. 0.18;
Matches 98; Conservative 60; Mismatches 184; Indels 93; Gaps 22;

QY 281 GTSVKNYSWFDY-----YLDQRLNSQAEGKRLLDV-FDVHVTPEAMGGGIRITNEVG 334
Db 93 GQTTDVTFESWTDSTDFKGYFVSVDAGSLGGQYSIDYSSDFAKYPR-----YGYISEFS 147
QY 335 NDETKKARQAPRTLMDPTYKEDSW-IAQWNSAFPLLPRLKQSVKYPGKTLALTEYS 393
Db 148 SNET-AAESAAKVNELAODYKINAWQFYDWMWRHETMIKRTGGTIDP-----TWIDLF--- 199
QY 394 YGENDISGGIAMTDVLGILGKNDVIMANYWKLKGANNYSAAKLYRNYD--GKNATF 451
Db 200 ---NRQISWPTINNOIAAIHNQGAAMA-----YAMIYAARENSYSGFVNPEW 244
QY 452 GDISVNAOTS---DIVNSVHVASVTDASYKELHLIVNKNKSMDSAFDAQFD-LSGETTYS 506
Db 245 GMYMDPAHTKOLDVDVFGNNSTYMYLFDPANAGWQOFIHEQYLDATQATANFDGIHIDOMGQ 304
QY 507 SGKIWGDKNSQIKAVAPITQISGNRTYTPPLTAVHIVLTADNDTPVPVDPEDPESFT 566
Db 305 RNNIYDYGNSIDL-----ATRET---PFKAATKLTAAANS----QDFMTFN 346
QY 567 LRAEAGDGKVDLSWDSAGSVGVSVQVATYENGPFPAASVNL-VETSYTDT-NVTNGTSY 624
Db 347 I---VDGTVN-GWAAND-----VSKNANVDLYSELWHLNS--SY 380
QY 625 Y---YKITAKTACTSESNVLKAVRTPVDGPDYAEADGTLKGITVSESSGFGSGT 680
Db 381 MOKDYIDSLRANSQ-NKAVVLAAYMYNGENIGDYEAEDAALQHTAVNTDHAGYTGSGF 439
QY 681 VTNFHNAGDSLMTMTI 695
Db 440 VDOFADVNDSVTFTI 454

RESULT 14
CAIC_MOUSE
ID CAIC_MOUSE STANDARD; PRT; 3067 AA.
AC Q60847;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=SWISS WEBSTER, AND C57BL/6J; TISSUE=Skin;
RX MEDLINE=9610761; PubMed=8601036;
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
RT "Primary structure of the long and short splice variants of mouse
RT collagen XII and their tissue-specific expression during embryonic
RT development.";
RL Dev. Dyn. 204:432-445(1995).
CC -!- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -!- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER

OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT ISOFORM CHAINS. ONLY THE LONG VARIANT IS A TROPOGLYCIN.

-1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHOONDRIUM, SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.

-1- DEVELOPMENTAL STAGE: LONG FORM OF COLLAGEN XII, XIIA, IS THE PREDOMINANT FORM AT EARLY STAGES (ED7 AND I1); AT LATER STAGES OF DEVELOPMENT (ED15 AND I7) THE SHORT FORM, XIIB, BECOMES THE MAJOR FORM. AS THE SHORT FORM BECOMES THE MAJOR PRODUCT, THE LONG SPLICE VARIANT CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.

-1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).

-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).

-1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.

-1- SIMILARITY: CONTAINS 4 VWFA DOMAINS.

-1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.

THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U25652; AAA9719.1; -;
HSP; P02751; IITF.
MGD; MG1:88448; Coll2al.
InterPro; IPR000087; Collagen.
InterPro; IPR003961; FN III.
InterPro; IPR003962; FnIII_repeat.
InterPro; IPR003129; TSPN.
InterPro; IPR002035; VWFA.
Pfam; PF01391; Collagen; 4.
Pfam; PF00041; fn3; 18.
Pfam; PF02210; TSPN; 1.
Pfam; PF00092; vwa; 4.
PRINTS; PR00014; FNTYPEIII.
PRINTS; PR00453; VWFADOMAIN.
SMART; SM00060; FN3; 16.
SMART; SM00210; TSPN; 1.
SMART; SM00327; VWA; 4.
PROSITE; PS02034; VWFA; 4.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.

SIGNAL 1 24
FT CHAIN 25 3067
FT COLLAGEN ALPHA 1(XII) CHAIN.
FT FIBRONECTIN TYPE-III 1.
FT VWFA 1.
FT FIBRONECTIN TYPE-III 2.
FT VWFA 2.
FT FIBRONECTIN TYPE-III 3.
FT FIBRONECTIN TYPE-III 4.
FT FIBRONECTIN TYPE-III 5.
FT FIBRONECTIN TYPE-III 6.
FT FIBRONECTIN TYPE-III 7.
FT FIBRONECTIN TYPE-III 8.
FT VWFA 3.
FT FIBRONECTIN TYPE-III 9.
FT FIBRONECTIN TYPE-III 10.
FT FIBRONECTIN TYPE-III 11.
FT FIBRONECTIN TYPE-III 12.
FT FIBRONECTIN TYPE-III 13.
FT FIBRONECTIN TYPE-III 14.
FT FIBRONECTIN TYPE-III 15.
FT FIBRONECTIN TYPE-III 16.
FT FIBRONECTIN TYPE-III 17.
FT FIBRONECTIN TYPE-III 18.
FT VWFA 4.
FT 2209
FT 2309
FT 2501

DOMAIN	2456	2751	NONHELICAL REGION (NC3).
DOMAIN	2752	2899	TRIPLE-HELICAL REGION (COL2) WITH 1 IMPERFECTION.
DOMAIN	2900	2942	NONHELICAL REGION (NC2).
DOMAIN	2943	3045	TRIPLE-HELICAL REGION (COL1) WITH 2 IMPERFECTIONS.
DOMAIN	3046	3064	NONHELICAL REGION (NC1).
DOMAIN	802	802	TO CHONDROITIN SULFATE (POTENTIAL).
BINDING	893	893	TO CHONDROITIN SULFATE (POTENTIAL).
BINDING	985	985	TO CHONDROITIN SULFATE (POTENTIAL).
SITE	866	868	CELL ATTACHMENT SITE (POTENTIAL).
SITE	2786	2786	CELL ATTACHMENT SITE (POTENTIAL).
SITE	2896	2898	CELL ATTACHMENT SITE (POTENTIAL).
MOD_RES	2945	2945	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2948	2948	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2951	2951	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2960	2960	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2966	2966	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2969	2969	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2972	2972	HYDROXYLATION (BY SIMILARITY).
MOD_RES	2984	2984	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3001	3001	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3004	3004	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3015	3015	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3024	3024	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3027	3027	HYDROXYLATION (BY SIMILARITY).
MOD_RES	3030	3030	HYDROXYLATION (BY SIMILARITY).
MOD_RES	869	872	POLY-THR.
CARBOHYD	704	704	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	1769	1769	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	2212	2212	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	2533	2533	N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD	2684	2684	N-LINKED (GLCNAC. .) (POTENTIAL).
VARSPLIC	25	1190	MISSING (IN SHORT ISOFORM).
SEQUENCE	3067 AA;	334094 MW;	3FB5DEFB8A2CDB95 CRC64;

Query Match 3.7%; Score 129.5; DB 1; Length 3067;
Best Local Similarity 21.5%; Pred. No. 1.6;
Matches 134; Conservative 61; Mismatches 232; Indels 197; Gaps 31;

QY	117	KAECEKPGAVTTSFHDQSLKLGAYSILVTLPMAGYVAKDNGSVQSEQAPSARWQVNA	176
DB	653	KAEWSPGDNVFSYH-----VTY-----KANGDDEVTTVVEPASSTSVLNS	694
QY	177	KNAPFOLQPD----LNDNOVYAEFVFLVKYGAAGTAKGVKGYALDNEPALWSHTPR	232
DB	695	-----LRPETLYLVNVTAEYED---GFSVPITGETT-AEVKGV-----PR	731
QY	233	IHGKVVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGWYWFV	292
DB	732	-----NLKVTDETTDSFKL-----TW---747	
QY	293	DYLDQMLNSQAEKRLLDVDFVHWYPEAMGGIRITNEVGNDETKKARMOAPRTLWDP	352
DB	748	-----SOAPGRVLR--YRIRYRP-VSGGESKESVTPANQRKLTLENTPTOKY--	792
QY	353	TYKEDSWIAQWNSAFPLPLRLKQSVDKYYPGTKL---ALTEYSYGGENDISGGIAMTDV	409
DB	793	---EISVIAEYPSG-----PCSPLTGNAATEVRGNPDLRVS DATIST	833
QY	410	LGIL-----GKNDVYMANVW-----KLKDGANNVYSAAYKLYR	442
DB	834	LKLSWSRAPKVKQYLVTYTPAAGGETQBTVRGDTTMTMLRKLKEG-TQYDLSVTALYA	892
QY	443	N-----YDCKNATFGD--ISVNAQTSIDVNSSVHVASVTDASYKELHLIVMKNMSDAPDA	495
DB	893	SGAGEALSCKGTLEERGSPQNLVTKDITDTSIGAYTWSAPGMVGRYRVSWKSLYDIEA	952
QY	496	QFDLSGETTYSCKTGWGDKNSQIKAVAPITQISGNRF-TYTV---PPLTAHYHVLTA	551
DB	953	-----GETTLG-----DAIHTMIENLOPEKTKYKISVFATYSSGEGEPTVG-DATTELS	1000

QY 552 NDTPTPPVPEDPESFTLRAEAGDKGVDLSWDASSG-VVGYSVQRTY--ENGPFAAVASNL 608
Db 1001 QSKILRVDDETEHMR-----VWKAAPGVVNY---RVYRPGGGROWAK-- 1046
QY 609 VETSYDT--NVINGTSYKKITAKTSGTSESNVLKAVRTPVDGPDRIYAEADGTLKG 665
Db 1047 VPPTVTSLKRLQPOTTYDITVLPYKTBGKLRQSGTTASRFKSPRNKLTSDPTMSS 1106
QY 666 TIVSESGTGFSGTGYVTFHNGD 689
Db 1107 FRVTWEPAPGEVKGKVTFFHTGD 1130

RESULT 15

ABFA_ASPNG STANDARD; PRT; 628 AA.
AC P42254;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alpha-L-arabinofuranosidase A precursor (EC 3.2.1.55) (Arabinosidase
A) (ABF A).
GN ABFA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-35 AND 63-77.

RC STRAIN=CBS 120.49 / N400;

RX MEDLINE=95093616; PubMed=8000538;

RA Flapphi M.J.A., Visser J., van der Veen P., de Graaff L.H.;

RT "Arabinase gene expression in Aspergillus niger: indications for

coordinated regulation.";

RL Microbiology 140:2673-2682(1994).

CC -!- FUNCTION: ACTS ONLY ON SMALL LINEAR 1,5-ALPHA-LINKED

L-ARABINOFURANOSYL OLIGOSACCHARIDES.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-

arabinofuranoside residues in alpha-L-arabinosides.

CC -!- PATHWAY: INVOLVED IN DEGRADATION OF THE PLANT CELL WALL

POLYSACCHARIDE L-ARABINAN.

CC -!- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L29005; AAC41644.1; -

KW EMBL; A27979; CAA01904.1; -

DR Hydrolyase; Glycosidase; Glycoprotein; Signal.

FT SIGNAL 1 25

FT CHAIN 26 628 ALPHA-L-ARABINOFURANOSIDASE A.

FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 610 610 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 628 AA; 67832 MW; 4F3E476FF93468AD CRC64;

Query Match

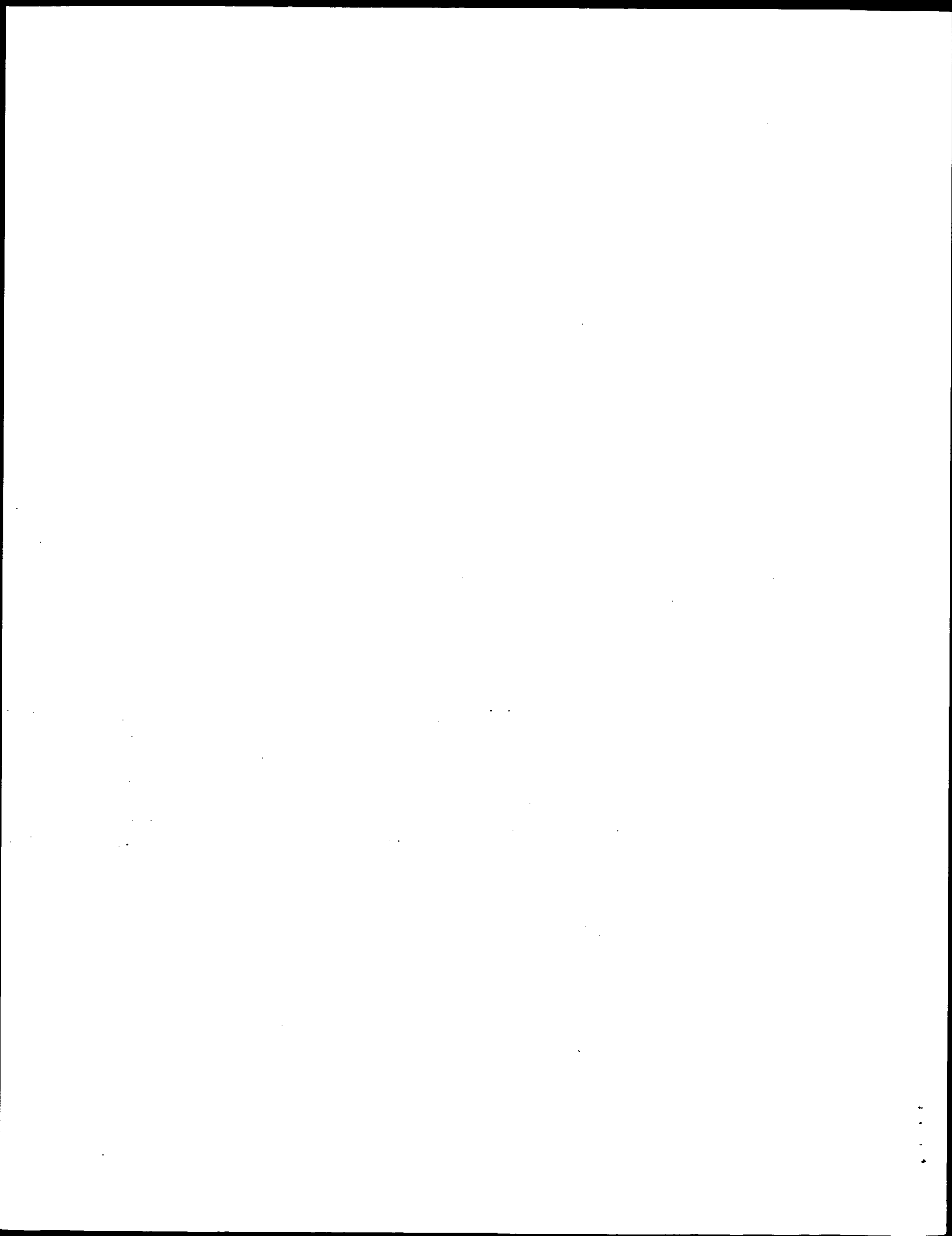
Best Local Similarity 3.7%; Score 128.5; DB 1; Length 628;

Matches 88; Conservative 53; Mismatches 123; Indels 149; Gaps 22;

QY 356 EDSWIAQWNSAFLLPRL-KQSDKYYPGTKLALTEYSY----- 394
Db 254 ENRW--KWNETIGDLCDRGREGTWTYYNTDGLGLHEFYFWCEDLGLVPLGVMDGFALE 311
QY 395 -GGENDISGGIAMT---DVLG---ILGKND-----VYMANWK 425
Db 312 SGGNTPLTGD-ALTPYIDVDVLEILGDTSTTYGAWRAANGOEFPNLTMYEIGNEDM 370
QY 426 LKDGANNYVSAAYKLYRNYDKGNATGDISVNAQTSIDVNSSVHASVTDASYKELHLIVM 485
Db 371 LGGGCESY---AERFTAFYDAIHAAYPDLILIASTSEA--DCLPESMPESGWVDYHDYST 425
QY 486 NKSWDAFAQDFDLSGETTYSSGKI-----W-----GFDKNSSQIK--AV 523
Db 426 PDGLVGQFYFDNLNRSVPYFIGEYSRWEIDWENMKGSVAEAFVFMIGFERNSDVVKMAAY 485
QY 524 APITQ-ISGNRFT---YTVPP-----LTAYHIVLTADNDTVPVPPVEDPESTLRAEAG 572
Db 486 APLLQLINSTQWTPDLIGYTQSPGDFLSTSYV-----OEMFS--RNRG 528
QY 573 DGKVDLSWDASSGVVGYSVQRTYENGPPAAVASNLVETSYTDTNVTNGTSYKKITAKT 632
Db 529 DTIKEVTSDS-----DFGLYVWASSA-----GDSYTVKVL---A 559
QY 633 KAGTSESNVLKAVRTPV-----DGPRIYAEADGTL-----KGTIVSSGT 673
Db 560 NYGSETQDLIVSIPGTSTGKLTVLADSDPDAYNSDTQTLVTPSESVQASNGT 612

Search completed: July 13, 2002, 00:59:34

Job time: 629 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:58:44 ; Search time 91.54 seconds
(without alignments)
1313.431 Million cell updates/sec

Title: US-09-784-554b-6
Perfect score: 3509
Sequence: 1 MKAKNSSNLSKRSKWLPPV.....SGTGYVNEHAGDSLMTI 695

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1465.5	41.8	1601	2 P71140	P71140 clostridium
2	1374	39.2	606	16 Q977Y3	Q977Y3 clostridium
3	1333	38.0	1751	2 Q9A0G4	Q9A0G4 caldicellul
4	967	27.6	632	2 Q52743	Q52743 ruminococcu
5	943	26.9	808	2 Q934F9	Q934F9 ruminococcu
6	515.5	14.7	209	2 Q9AQG9	Q9AQG9 caldicellul
7	221.5	6.3	997	2 Q924I1	Q924I1 bacillus sp
8	161.5	4.6	1090	16 Q928J2	Q928J2 listeria in
9	161	4.6	1053	2 P77865	P77865 fibrobacter
10	161	4.6	1864	2 Q03658	Q03658 unidentified
11	157	4.5	717	2 Q9KHB3	Q9KHB3 bacillus ci
12	157	4.5	1042	16 Q97GG3	Q97GG3 clostridium
13	155	4.4	1053	2 P71329	P71329 fibrobacter
14	148	4.2	2032	2 Q45643	Q45643 bacillus sp
15	146.5	4.2	1282	2 Q46348	Q46348 cytophaga s
16	144.5	4.1	1433	2 Q9L040	Q9L040 streptomyce

17	141.5	4.0	1433	2	P75020	P75020 streptomyce
18	140	4.0	881	2	Q9AF09	Q9AF09 pseudomonas
19	137.5	3.9	699	2	Q48494	Q48494 kurthia zop
20	137	3.9	492	2	Q52155	Q52155 burkholderi
21	136	3.9	2167	16	Q92EK5	Q92EK5 listeria in
22	135	3.8	7107	5	Q9V4F7	Q9V4F7 drosophila
23	134.5	3.8	26926	4	O10466	O10466 homo sapien
24	133.5	3.8	6831	5	Q23550	Q23550 caenorhabdi
25	133.5	3.8	7160	5	Q23551	Q23551 caenorhabdi
26	132.5	3.8	797	2	Q93IE7	Q93IE7 bacillus sp
27	132	3.8	583	2	Q45822	Q45822 clostridium
28	132	3.8	628	3	Q96X54	Q96X54 aspergillus
29	131	3.7	157	1	P94805	P94805 haloferax s
30	130	3.7	700	2	Q30678	Q30678 xanthomonas
31	130	3.7	2018	2	Q9E224	Q9E224 bacillus st
32	130	3.7	6875	6	Q28733	Q28733 oryctolagus
33	128.5	3.7	824	16	Q9AB96	Q9AB96 caulobacter
34	128.5	3.7	1441	16	Q9CFL1	Q9CFL1 lactococcus
35	128.5	3.7	1578	16	Q92E25	Q92E25 listeria in
36	128	3.6	1915	2	Q9RPL0	Q9RPL0 acetivibrio
37	127	3.6	817	13	Q07784	Q07784 gallus gall
38	124.5	3.5	993	2	P96992	P96992 streptomyce
39	124.5	3.5	1063	16	Q9KD05	Q9KD05 bacillus ha
40	124.5	3.5	1386	16	Q92DL0	Q92DL0 listeria in
41	124	3.5	495	2	Q3LBN6	Q3LBN6 leptothrix
42	124	3.5	967	3	Q08294	Q08294 saccharomyc
43	124	3.5	1001	3	Q05164	Q05164 saccharomyc
44	124	3.5	1826	16	Q98727	Q98727 rhizobium l
45	123	3.5	492	2	Q9XDX7	Q9XDX7 burkholderi

ALIGNMENTS

RESULT 1

P71140 PRELIMINARY; PRT; 1601 AA.
AC P71140;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ENDOGLUCANASE.
GN CELJ.
OS Clostridium thermoceum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fl;
RX MEDLINE=96422002; PubMed=8824619;
RA Ahsan M.M., Kimura T., Karita S., Sakka K., Ohmiya K.;
RT "Cloning, DNA sequencing, and expression of the gene encoding
RT clostridium thermoceum cellulase celJ, the largest catalytic
RT component of the cellulosome."
RL J. Bacteriol. 178:5732-5740(1996).
DR EMBL: D83704; BAA12070.1; -
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR004197; celD_N.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000601; PKD_domain.
DR Pfam; PF02927; celD_N_1.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00759; Glyco_hydro_9; 3.
DR Pfam; PF00801; PKD; 1.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 1.
DR PROSITE; PS50093; PKD; 1.
SQ SEQUENCE 1601 AA; 178058 MW; 31E85D77F8642565 CRC64;


```

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CELE.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOK7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078042; AAK06394.1; -
DR HSSP: Q06851; INSC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR005666; Lipocln_cytFABP.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR ProDom: PD001947; CBD_3; 3.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 38.0%; Score 1333; DB 2; Length 1751;
Best Local Similarity 48.5%; Pred. No. 2.4e-77;
Matches 262; Conservative 73; Mismatches 179; Indels 26; Gaps 7;

QY 32 PAPTIVHGGT-----AKVTIKVDTSKDKRKPISPYIYGTQDLAGDENLAARLGGNRT 86
DB 1215 PPTVTVTPPTPAVTPDVKISDTISRGRTKISPIYIGANDIQGVH-PARRLGGNRLT 1273
QY 87 GYNWNNNSAGSDMQQSDNFGGLTKAECEKPGAVTTSFHDQSLKLGATSLVTL 146
DB 1274 GYNWNNNSAGSDWYHSSDDIMVIMGITGNDKNVPAVYKFEHSIKQNAVSAITLQ 1333
QY 147 MAGYVAKDNGSVQSEQAPSARWVNVNAKNAPFQLPDLNDQVYADEFVNLVKY 206
DB 1334 MVGYVAKDNGTGVSEETAPSPRAEYKFKDGLSLQPDVNDVYVWDFEINTLINKY 1393
QY 207 AASTAGYKGYALDNEPALWSTHPRHGEKVKYXXXXXXKXXXXXXXXXXXXX 266
DB 1394 RSSATGKGYLDNEPDLWFTTHPRHPRKVTGSELINKSVELAKVIKTLDPDAEIFGP 1453
QY 267 XXXXXXXXFKLHXI--GTSVKGYSWFVDYVYLDQMLNSQAEGRLLDVEDVHWYPEAMG 324
DB 1454 ASYGFVGYLTLQDAPDNQVGNHWFWSWLEQMKKASDSFGKRLDVLVDIHWYPEAQV 1513
QY 325 GGIRITNEVGNDETKK----ARMQAPRLWDPYK-----EDSWIAQWNSAFPL 370
DB 1514 GGVRICTD-GENSTSRDVAIARMQAPRLWDPYKTKQGITAGENSINQWPEYVPL 1572
QY 371 LPRLKQSDVKYPGTKKALTEYSYGGENDISGGIAMDVLGILKNDVYVANYWKLK 430
DB 1573 LPNIKADIKYPGTKKLAITEFDYGGKDHISGGIALADVLGIFGKYGYMAARW---GDS 1629
QY 431 NNYSAAVKLYRNDGKNATGDISVNAQTSIDVNSSVSHASVTDASKEHLHIVNKSMD 490
DB 1630 GSYAAQATNLYNDGKSGRSGTCSVAETDVENMPVIASIEGEDDSTVHIILINRYD 1689
QY 491 SAFDAQFDL-SGETTYSKGKINGFDKNSQIKAVAPITQISGNRTYTPVPLTAVHIVLT 549
DB 1690 RKLKAEIKWNTYVTGGEIYGFDSSTSQIKMGVLNIGNQNTTIEVPNLTVTHIVLTS 1749

Query Match 27.6%; Score 967; DB 2; Length 632;
Best Local Similarity 33.2%; Pred. No. 2.5e-54;
Matches 225; Conservative 102; Mismatches 246; Indels 104; Gaps 17;

QY 23 CTIIVG-GALPAPTIVHGGTAKVTIKVDTSKDKRKPISPYIYGTQDLAGDENLAAR 78
DB 16 CTLSAPLGLVP-ETKVNAGGFDWNIKVDLKGKKEISPLIYGVNQVTTDLKSVKTTAVR 74
QY 79 RLGNRMTGYNWNNNSAGSDMQQSDNFGGLTKAECEKPGAVTTSFHDQSLKLG 138
DB 75 Q-GGNRTATNWNNSAGSDWYHSSDDIMVIMGITGNDKNVPAVYKFEHSIKQNAVSAITLQ 125
QY 139 A-YSLVTLPLMAGYVAKDNGSVQSEQAPSARWVNVNAKNAPFQLPDLNDQVYADEF 197
DB 126 VDKYKMTTLQAGYVSADKDGTVKEDEVAPSKRNEVKFTKGAFDEPDLTDGVVYMD 185
QY 198 VNFVKYKGAASFTAGVKGYALDNEPALWSTHPRHGEKVKYXXXXXXKXXXXXXXXXXXXX 257
DB 186 VNYIINKLGDQSPTGIQGYSLDNEPVLWNTHPRVHPPEVTTIEELGNKSIELAKAVKKL 245
QY 258 XXXXXXXXKXXXXXXXXKFL-----HXIGTSV--KGNYSWFVDYVYLDQMLNSQAEGRKL 310
DB 246 DPKAEIFGPALYGYTAPDHLDDDEQHTESGDVKSNNYHWLDCYLDQMKKASEEGTRL 305
QY 311 LDVFDVHWYPEAMGGGIRITNEVGNDETKKARMQAPRLWDPYKEDSWIAQWNSAFPL 370
DB 306 LDVLDIHWYSESARTG-----AEDRVQSVRTLYEEGFSNLSIGQWCMQNPV 353
QY 371 LPRLKQSDVKYPGTKKALTEYSYGGENDISGGIAMDVLGILKNDVYVANYWKLK 430
DB 354 LPTIKASIDITYPGTKKLAITEYNFKGEDTSGTIAQAEALGCFADQGVYATLW----GG 409
QY 431 NNYSAAVKLYRNDGKNATGDISVNAQTSIDVNSSVSHASVTDASKEHLHIVNKSMD 490
DB 410 EPIIISGINLYTNDGKGGCFGTDLTPASTEDVSKSTYAAVNDGDESQVTVITNKNMT 469
QY 491 SAFDAQFDL-SGETTYSKGKINGFDKNSQIKAVAPITQISGNRTYTPVPLTAVHIVLT 549
DB 470 EAENAVIDLKNSKDYKSAAVYAVYVGNDOVRLDLVDKDKNVNVELPASAAVYV 529
QY 550 ADNDTPVPVPEDEPSTTLRAEAGD--KVDLSWDASSGVVGYSVQRATYENGPFPAVASNL 608
DB 530 DD-----AAAFDGEKI-----YEEKVTEKTEEFKDPSSMI 560
QY 609 VETSITDNTVNTGTSYKYYITAKTAKTSESNNVLKAVPTPVDGDPDRPDAEDGTLKGTIV 668

```


Db 178 RMGVLNSIQNNNTTITIEVPLNTVYHIVLTS 207

RESULT 7
Q92411 ID Q92411 PRELIMINARY; PRT; 997 AA.
AC Q92411; 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DE CELLULOSE PRECURSOR (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. BP-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
from *Paenibacillus* sp BP-23: evaluation of its performance in pulp
refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AJ133614; CAB38941.1; --
DR HSSP; P26221; 1TF4
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000566; Lipocin_cytFABP.
DR Pfam; PF00942; CBD_3; 2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR Prodom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL.
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match 6.3%; Score 221.5; DB 2; Length 997;
Best Local Similarity 23.9%; Pred. No. 6.1e-06;
Matches 98; Conservative 63; Mismatches 130; Indels 119; Gaps 20;

QY 379 DKY-----YPTGKIALTEYSYGGENDISGGIAMTDVLGILGKNDVYVANYWKLK-DGAN 431
Db 452 DQYTDIDIGYVSNEVA-TDYNAG-----FTGALAKNLL--YQNHQPLANPAPVKGDE 504
QY 432 NTVSAAYK-----LYRNYDGKNATFGD-----ISVNAQTSDI--V 464
Db 505 YVEEAAVRSSGSNYTEIRALLNNRSGWPARMGDQLSFYFLDLSEVAAGRTVSDVQVTV 564
QY 465 NSSVHASVT----DASYKEHLIWNKSMDSAP-----DAQFDLSETTYSGSKI 510
Db 565 SSSEGAIVSQPVVDAA-KRIAYITANFSNTKIYPGGEGNYKEVQFRITGPQ-----GA 618
QY 511 WGFDRNSSQIAKAVAPITQISNRFTYTPPLTAYHI-----VLTADNDTPVPVPEDES 564
Db 619 WN-----FANDPSYQNL-TGNPKVSNVPIPVYDAGVKVSGEPGVTPVAVPAA 665
QY 565 -FTLRAEAGDKVLSWADSSGVGVYQVQRYATYENGPPAAVSNLVTSTYTDNTVNGTS 623
Db 666 PAGVQAVAGNSQVALNWSAGAVSYTVKRAEVSQGGPYTTTAAAGVNGLTNTGLTNGKT 725
QY 624 YYYKLTAKTAKTSSESN-LKAVPR-----TPVDPDRYEA 658
Db 725 YYYVTVAVNSAGESPASVQVSGMPQAAATVPVAVTSLGTAQNNQNLNLSWTAAGSASYTV 785

QY 659 EDGTLKGTIVE-----SSCTGFGTCY----VTNFHAGDSLTFWTI 695
Db 786 QRAVAGGYTDTVATGLAVLNTYNDTALNGTSYSYRAAVNASGQTLNNV 835

RESULT 8
Q928J2 ID Q928J2 PRELIMINARY; PRT; 1090 AA.
AC Q928J2;
DT 01-DEC-2001 (TREMREL. 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
DE LIN2540 PROTEIN.
GN LIN2540.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkak G.,
RA Madoeno E., Maltournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of *Listeria* species.";
RL Science 294:849-852(2001).
DR EMBL; AL596172; CAC97767.1; --
DR L2stList; LIN2540; --
KW Complete proteome.
SQ SEQUENCE 1090 AA; 121745 MW; 5294EF5A0327ECBA CRC64;

Query Match 4.6%; Score 161.5; DB 16; Length 1090;
Best Local Similarity 18.2%; Pred. No. 0.051;
Matches 146; Conservative 91; Mismatches 272; Indels 293; Gaps 33;

QY 39 GQAKTVTIKVDTSKDRKIPSPYIYGTNQDLADENLAARRLGNRMRTGYNNWENN----- 93
Db 355 GDMTNMLDYVVISGKDQND---IVNNYTDITGKTTLLPKWAFGLMWSANEDRESVDST 410
QY 94 -MSNAGSD-----WQSSDNFLCNGGLT-----KACEKPGAV 126
Db 411 ALSNAKENIIPATGFCVLEONSDEETIYNWDATYATKKNGEFFAYDDFTNGKWTNPKAM 470
QY 127 TTSFHDOSLKLGYSLVTLTP-----MAGYVAKDNGSVQSEQAPSAR 169
Db 471 VDSVHDAGMIVLWQVPLKDDNTVVEQRDNDEYMWISQYSADDGTGA---PYRVPASQ 527
QY 170 WNOVVNAKNAPFOLQPLDNDNQYAD-----EFVNFVLKYGAASTKAGVKGYALDNEPAL 225
Db 528 W-----FCNGILLDFTNKDAVNNWTSQREYLLTDVGDIDGPKTDGEMV 570
QY 226 WSHTHPHIHEKVKYXXXXXXXKXXXXXXXXXXXXXXXXXXXXXFKLHIGTSVK 285
Db 571 WGRDITTFNGEK-----GOEMR 587
QY 286 GNT-SWFVDYIYLDQMR-LNSQA-----EGKRLLDVDFVHWYPEAMGGGIRITNEVGNDET 338
Db 588 NRYPTDYVSSYFDFAKSINPEAVSFRSGTSGAKKSGIYH-----SGDQT 632
QY 339 K-----KARMQAPRT-----LWD-----PT-----YKEDSWIAQWNSAFLPLPR 373
Db 633 STEDSFQASVKAGLSASTSGSVYAWMDMAGFTGNYPTAEYKRTATA---AFAPIMQF 688

```

Qy 374 LKGSVDYKYPGTKLALTEKSYGGENDISGIAWTDVILGKNDVYMAN-----YWK 425
Db 689 HSEKSDP-----SPSEERSPWNAVARTGDETILPTFKYLYTRMNLPIYTA 736
Qy 426 LKDGANNVSAAYKLYRNYD-----GKNATFGD-----LSVNAQTS----- 461
Db 737 AKOTADNGKSMRQMAWDYPEDINTRNLDDQYMGFDLLVAPVQBGQTEKEVYLPGEW 796
Qy 462 -DIVNSSVHASVTDASY-----KELHLIVMN-----KSMDSAFDAQFDL-- 499
Db 797 IDIWNNGIHPGGETISYADVDPVLPFAKAGAIIPNNMTDGYOLGQVNGDLKAYDNLTF 856
Qy 500 -----SGETYSKINGFDKNSOIKAVAPITQISGNRFTYVPPITAYHIVTADNDTP 555
Db 857 RVPSGNSEYS----FYDDVNGGEMRDISVSEDFANEKVTNLPAM-----ADETTM 904
Qy 556 VPPVEDPESLTLR-AE-AGDGKVDLSDWASSGVVGVSVORATYENGPFPAVASNLVETSY 613
Db 905 QVFTEPTSVTVAGAEEAKADTLDAFNEATSAIYYDTVQNLTY-----IKTAA 952
Qy 614 TDTNVTNGTSYIYKITAKTAKTAGESNVLKAVPRTVPDGPDRYAEADGTILKGTIVESSGT 673
Db 953 TD-----ITQAILVNGVNHAP-----YEAEGHLTNVYSTASDHA 986
Qy 674 GFGSGTGVTFNHNAGDSLWTMI 695
Db 987 GYTGTGEVAGFDEEKAVEFDI 1008

RESULT 9
P77865 PRELIMINARY; PRT; 1053 AA.
AC P77865;
DT 01-FEB-1997 (TremBLrel. 02, Created)
DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)
DE ENDOGLUCANASE F (EC 3.2.1.4).
GN CELF.
OS Fibroblast succinogenes (Bacteroides succinogenes).
OC Bacteria; Fibroblast/Acidobacteria group; Fibroblast group;
OC Fibroblast.
OX NCBI_TaxID=833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97316456; PubMed=9172367;
RA Malburg L.M. Jr., Forsberg C.W.;
RT "Fibroblast succinogenes S85 possesses at least nine distinct
glucanase genes.";
RL Can. J. Microbiol. 39:882-891(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97316456; PubMed=9172367;
RA Malburg S.R., Malburg L.M. Jr., Liu T., Iyo A.H., Forsberg C.W.;
RT "Catalytic properties of the cellulose-binding endoglucanase F from
Fibroblast succinogenes S85.";
RL Appl. Environ. Microbiol. 63:2449-2453(1997).
DR EMBL; U39070; AAC45377.1; -.
FT CHAIN 24 1019 ENDOGLUCANASE F.
SQ SEQUENCE 1053 AA; 118586 MW; 37E8FC2AD9C6F6D5 CRC64;

```

```

Query Match 4.68; Score 161; DB 2; Length 1053;
Best Local Similarity 18.2%; Pred. No. 0.053;
Matches 99; Conservative 81; Mismatches 157; Indels 206; Gaps 25;

Qy 43 KTVIKYDTSKDRKPISPYI-----YGTNQ-----DLADENLA-----ARRL----- 80
Db 620 RPAVKLVKGEKVDLPNIPISGGVFLNALWDGMDLNDKFNKQVREYAKRNHGLIIRY 679
Qy 81 -GGNRMTGYNNENNSAGSDWQSSDNFLCNNGGLTKABCEKPGAVTTSFHQSLKGLA 139
Db 680 PGLRADDDHKEILDN-HDWMYDTDFL-----EWLKKTGS 715

```

```

Qy 140 YSLVTLPWAGYVAKDNGSGVQSEQAPSRNQVNVNAKNAPFQLOPDLNDNQVYADEFVN 199
Db 716 NAMFTVNF-----GSGTVKE-----AADW----- 734
Qy 200 FLVKKYGAASTKAGVGYALDNEPALWSHTHP-----RIHGEKVVXXXXXXXXXX 249
Db 735 ---VXHTNIDRKAGLLYWEIGNE---IYGNWHPYKYEYKDGDTIYGR----- 777
Qy 250 XXXXXXXXXXXXXXXXXXXXXXXXFKLHXICTSYKGNYSWFVDYIYLDQMLNSQAEGR 309
Db 778 -----ARKIEAMKKVDPPIKVALGV-LEGDMN---DKVL-----AETGD 814
Qy 310 LLDVFDVWHPMAMGGGIRITNEVNDKTKARMAQPTL-----WDPTTKE 356
Db 815 IADGLVLIHHPQHF-----EENDFAMLSAPQTLTAIYERLHKVVDKWTTKYNK 863
Qy 357 DS---WIAOWNSAFPLPLPRLKQSDVKYVPGTKLALTEYSYGGENDISGGIAMTDLGI 412
Db 864 DKKIELWTEWN-----SVD-FNPGQPL-----SYENGLFVADYLCM 900
Qy 413 LGKNQVIMANYWKL-----KDGANNVSAAYKLYRNYDGKNATEGDISVNAQTSDIYNS 466
Db 901 LATENVNAQYWDIHDITPEGGDYGLTRSGECCMCPRPSYAWAFOMASDALRGKLMKT 960
Qy 467 SVHASVTD-----SYKELHLIVMKNKSMDSAFDAQFDLSGETTYSKINGFDKNS 518
Db 961 TIKGD-EDALLTAYTWNKNGKQLLVNKSYPSEFDIKLDIPG---FKGAKVQTLDKTSE 1017
Qy 519 QIK 521
Db 1018 KIK 1020

RESULT 10
Q03658 PRELIMINARY; PRT; 1684 AA.
AC Q03658;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DE EXO-MALTOPEPTAHOYDROLASE PRECURSOR (EC 3.2.1.-) (AMYLASE).
GN AMY180.
OS unidentified bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=23338;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-62.
RC STRAIN=UG0163/26;
RX MEDLINE=90336627; PubMed=1696201;
RA Candussio A., Schmid G., Boeck A.;
RT "Biochemical and genetic analysis of a maltopentaose-producing amylase
from an alkaliphilic gram-positive bacterium.";
RL Eur. J. Biochem. 191:177-185(1990).
CC -!- FUNCTION: CLEAVAGE OF STARCH TO LINEAR MALTOPEPTAOS.
CC -!- SIMILARITY: STRONG. TO ALPHA-AMYLASES AND TO THE BETA-AMYLASES OF
BACILLUS POLYMYXA AND BACILLUS CIRCULANS.
DR EMBL; X53373; CAA37453.1; -.
DR HSSP; P02751; ITTF.
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF00041; fn3; 3.
DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN_1.
KW Hydroxylase; Glycosidase; Polysaccharide degradation; Signal; Repeat.
FT SIGNAL 1 37
FT CHAIN 38 1684 EXO-MALTOPEPTAHOYDROLASE.
FT DOMAIN 180 441 REPEAT-RICH REGION.
FT DOMAIN 977 1233 REPEAT-RICH REGION.
FT REPEAT 180 185 REPEAT 1.
FT SIMILAR 180 441 STRONG, WITH ALPHA-AMYLASES.
FT REPEAT 329 340 REPEAT 2.

```


FT REPEAT 381 384 REPEAT 3.
FT REPEAT 435 441 REPEAT 4.
FT REPEAT 977 1026 REPEAT A (APPROXIMATE).
FT SIMILAR 977 1233 STRONG, WITH BACTERIAL BETA-AMYLASES.
FT REPEAT 1080 1129 REPEAT A (APPROXIMATE).
FT REPEAT 1184 1233 REPEAT A (APPROXIMATE).
SQ SEQUENCE 1684 AA; 186497 MW; C151079BF0DD2B35 CRC64;

Query Match 4.6%; Score 161; DB 2; Length 1684;
Best Local Similarity 20.6%; Pred. No. 0.1;
Matches 146; Conservative 62; Mismatches 256; Indels 244; Gaps 29;

QY 64 GTNQDLAENLAARLGG-----GNRTGTYNNENNNSNAGSDWQ---QSSDNFLCNN 112
DB 461 GGVQVYFGDE--TARPLDGGSDPQGTGR--SSMNWANTNQVLSHWKQLGQFRNNHIAIG 517
QY 113 GGLTKAECEKPGAVTTSFHQSLKLGAYSLVTLPMAGYVAKDNGSGVQSEQAFAFSARWQ 172
DB 518 AGAHQKLSDSPTFEARTVESDDI-----VDEVVATGAQGT 554
QY 173 VYNAKNAPQLQDLNDNQYADEFV--NFLVKYKGAATKAGVKGYALDNEPALWSH 230
DB 555 AVTVEGV-----FEDGTWVRDAYTGDDETTVTK--GTATFTAGTQGIILIENTA----- 600
QY 231 PRHGEKVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXFKLHXIGTSVKGNYSW 290
DB 601 -----EPVTNL-----PIVSATPGNSSF 618
QY 291 FVDYLDQMLNSOAEGRKRLDVFVHWYPEAMGG--GIRITNEVGNDETAKRMOA----- 345
DB 619 RTDDITITLNVDRADWKYTLDSGDPADGLTFMDGEEIVIGADMEFDEATLRLYAENEN 678
QY 346 -----PRTLMDPTKESWTAQWNSAFLLPLRLKQSDVKYYPGT----- 385
DB 679 GIRTSYTYRKVDPPDALLEVEFKP---ADMGTPHI-----YYDTPEPEPEV 723
QY 386 -----KLATE---YSYGGENDISGGTAMTDVLG--ILGKNDV-----YMANYSWK 425
DB 724 TWTAPETMLVEDDQVYVYFENAEANLIFKDSKQIPGNEPFFDQIGWYDGVKWL 783
QY 426 LKDGANNYSAAKYLYRNDGKNATFGDISVNAQTS-----DIVNSSVHASV 472
DB 784 DSDPFREPKEPATPKNLSVYVNTETTVTFEWDQSDGVVVEYELRDDEDVVASITRTTF 843
QY 473 TDASYKELHLIVMNSKMSDAFDAQFDLSGETTYSYSGKIFGFKNSQIKAVAPITQISGN 532
DB 844 TDE-----DLNPDFTY-----TYSVAVGEGQKSA- 869
QY 533 RFTYTPPLTAYHIVLTADNDTPVPVDEPESFTLRAEAGDGKVDLSWDASSG-VVGYSV 591
DB 870 -----PSEALKVITLENDEKPEAPENLRI-ADITDTVTIINNNSNGYVTGTEV 921
QY 592 QRATYENGPPAAVASNLVTSYTDNTVNGTSYTYKITAKTKAGTSSESNVLKAVRTPVD 651
DB 922 LRD-----GVVIGETTRTTFIDTGLDADRTYTTIVALDGG-QKSDPSEALEVTQOE 973
QY 652 GPD-----RYAEDGTLKGTIV-----ESSGTGFS 676
DB 974 KPEGNLVTIYKKGFDTPYMHYRPEGG--EWTIVPGRMESEIAGYS 1019

RESULT 11
Q9KHB3
ID Q9KHB3 PRELIMINARY; PRT; 717 AA.
AC Q9KHB3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CHITINASE.
GN CHIL.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
RN NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-2;
RA Wang Y.L., Wang H.Y., Qin M., Zhang Y.Z.;
RT "Nucleotide sequence of a chitinase gene (chil) from Bacillus
RT circulans C-2";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AF265220; AAF74782.1; -;
DR HSP; P02751; ITTF.
DR InterPro; IPR001579; Chitinase.2.
DR InterPro; IPR003610; Chitin_bind3.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF02839; CBD_7; 1.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00495; ChCBD3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
KW Hydrolase; Repeat.
SQ SEQUENCE 717 AA; 76118 MW; BB0B019CDE72C198 CRC64;

Query Match 4.5%; Score 157; DB 2; Length 717;
Best Local Similarity 23.4%; Pred. No. 0.056;
Matches 80; Conservative 48; Mismatches 138; Indels 76; Gaps 15;

QY 356 EDSHIAOWN-SAFILPLRLKQSVKYPGKTKALTYSGGENDISGGIAMDVLGILG 414
DB 382 KNGYTRYNDTAKYPPFL-----YNASNKRFI---SYDDAESIGHKTAIKSKGLG 429
QY 415 KNDVYMANYSKLDGANNYSAAKYLYRNDG-----KNATFGDISVNAQTSDIVNSSVHA 470
DB 430 -----AMFWELSDGRNKTQNLKLSDLSTGGTVPPTDTPAPSVPGNARSTGYTASSV-- 481
QY 471 SVTDASYKELHLIVMNSKMSDAFDAQFDLSGETTYS--SGKIMGPKNSQIKAVAPITQI 529
DB 482 -----TLAWNASTDNV-----GVGTGYVINGTSVTSVTGTTATISGLAP---- 521
QY 530 SGNRFTYTPPLTAYHIVLTADN-----DTPVPVDEPESFTLRAEAGDGKVD 577
DB 522 -GTSYTFTRAKDAAGNLASAASNLSTVTVQPGDTPQPTV---PTNLTSTAKT-SSIT 577
QY 578 LSWDASS---GVVGYSVQRATYENGPPAAVASNLVTSYTDNTVNGTSYTYKITAKTKA 634
DB 578 LSWAASDNDVGVIGYEVNGT-----ALVTVSGTSAITVGLTADTSYTFVAKADAA 630
QY 635 G-----TSESNNLKVAPRTPVDGPDREAEADGTLKGTIVESSG 672
DB 631 GNLSAASALTAVKTEGVGTNPVSAWQANTAYVVGQLVTYNG 672

RESULT 12
Q97G63
ID Q97G63 PRELIMINARY; PRT; 1042 AA.
AC Q97G63;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SPOID-LIKE DOMAIN CONTAINING PROTEIN, PEPTIDOLYCAN-BINDING
DE DOMAIN.
GN CAC2506.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
RN NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 09:40:13 ; Search time 18262.7 Seconds
(without alignments)
4651.057 Million cell updates/sec

Title: US-09-784-554B-1
Perfect score: 4059
Sequence: 1 atgaggcgcaaaatagtag.....ccgagcgacagctaataa 4059

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	602.2	14.8	2977	6	A28170	A28170 B.lautus st
2	601.4	14.8	1854	6	AR012245	AR012245 Sequence
3	601.4	14.8	1854	6	I14943	I14943 Sequence 28
4	601.4	14.8	1854	6	I73663	I73663 Sequence 28
5	599	14.8	2831	1	BACCELAA	M76588 Bacillus la
6	511.2	12.6	6465	1	D83704	D83704 Clostridium
7	455.8	11.2	13603	1	AE007607	AE007607 Clostridi
8	413.8	10.2	4977	1	CDCMANA	L01257 Caldocellum
9	405.2	10.0	6005	1	AF078038S5	AF078042 Caldicell
10	405.2	10.0	6416	6	E35100	E35100 Truncated c
11	344.8	8.5	1423	1	AF013989	AF013989 Dictyoglo
12	268.6	6.6	3900	1	U39812	U39812 Caldicellul
13	193.4	4.8	4831	1	MRMNMAG	X90947 R.marinus m
14	187	4.6	2492	1	RFU08621	U08621 Ruminococcu
15	178.6	4.4	2692	1	REFL298117	AJ298117 Ruminococ
16	172.6	4.3	4567	1	AF163837	AF163837 Caldibaci
17	171.2	4.2	3237	1	AF200304	AF200304 Caldibaci
18	157.6	3.9	2201	1	AB044406	AB044406 Clostridi
19	140.6	3.5	1855	1	AF078038S2	AF078039 Caldicell
20	131.4	3.2	4161	1	BSP133614	AJ133614 Bacillus
21	128.8	3.2	2015	1	CTH242666	AJ242666 Clostridi
22	121.2	3.0	2020	8	PSE14BMAN	X97408 Piromyces s
23	119.6	2.9	2336	8	PSMANAGEN	X91857 Piromyces s
24	112.6	2.8	3120	1	CSCELYZ	Z69359 C.stercorari
25	103.8	2.6	3180	1	AY039744	AY039744 Bacillus
26	102.2	2.5	3180	1	AF078737	AF078737 Caldicell
27	93.2	2.3	11707	6	E35099	E35099 Truncated c
28	90	2.2	5437	1	CASR69XN2	AF036924 Caldicell
29	89.4	2.2	3262	1	AF078038S1	AF078038 Caldicell
30	88.8	2.2	2587	1	AB016164	AB016164 Bacillus
31	88.8	2.2	4241	1	CSCELB	X13602 Caldocellum
32	87.8	2.2	2029	6	E35142	E35142 Truncated c
33	87.8	2.2	2029	6	E35143	E35143 Truncated c
34	87.8	2.2	2029	6	E35143	E35143 Truncated c
35	87.4	2.2	1920	1	BACGLUB	M16195 B.subtilis
36	87.4	2.2	2589	1	AY044252	AY044252 Bacillus
37	87.2	2.1	5439	1	CDCCELA	L32742 Caldocellum
38	87	2.1	2175	1	BACCELD	M28332 B.subtilis
39	87	2.1	2175	6	E05425	E05425 DNA sequenc
40	87	2.1	6555	1	AF155197	AF155197 Acetivibr
41	86.4	2.1	5513	1	AT286105	Z86105 A.thermophi
42	85.8	2.1	1523	1	AF355629	AF355629 Bacillus
43	85.8	2.1	1800	1	BSEGLSG	Z29076 B.subtilis
44	85.8	2.1	2314	1	BS14GLUC	X67044 B.subtilis
45	85.8	2.1	2435	1	BSBGLUC2	X04689 Bacillus su

ALIGNMENTS

RESULT 1	A28170	A28170	2977 bp	DNA	linear	PAT 25-MAY-1995
LOCUS	A28170	B.lautus strain NCIM 40250	endo cellulase gene 1 (Endo1).			
DEFINITION	A28170					
ACCESSION	A28170					
VERSION	A28170.1	GI:905293				
KEYWORDS	Paenibacillus lautus.					
SOURCE	Paenibacillus lautus					
ORGANISM	Paenibacillus lautus					
REFERENCE	1	(bases 1 to 2977)				
AUTHORS	AN ENZYME EXHIBITING CELLULOSE ACTIVITY					
TITLE	Patent: WO 9110732-A 1 25-JUL-1991;					
JOURNAL	Location/Qualifiers					
FEATURES	source					
	1..2977					
	/organism="Paenibacillus lautus"					
	/db_xref="taxon:1401"					
	677..2779					
	/codon_start=1					
	/transl_table=11					
	/product="endo cellulase 1"					
CDS						

/protein_id="CAA01933.1"
 /db_xref="GI:905294"
 /translation="MKTRKRLFVSAALAVSLTMTVMPASVNAASDVFTTINTOS
 ERAATSPNITGTNODLSGTENWSSRLGNRLTGYWNENASSAGRWLHYSDDFLCG
 NGGVPDTCDFKPGAVVTAFAHDKSLENGAYSIVTLQAGIYVSRDKNGPVDSETPSPR
 WDKFEAFNAKPELQDNDQVYDEVEFLVNRVGNASTGKIKAYSLDNEPALMS
 ETHPRHPEQLQAELVAKSIDLSKAVKNDVPHAEIIPGALYGFAYLSLQDAPDWS
 LQGNYSWFIIDYLOOMKNAHQNGKRLDLDVHWYPEAGGGQRIYVFGAGNIDTOK
 ARVQAPRLSDPAPVOEDSWITFESSYPLPKLOSIQYTPGKLAITEFSGGDN
 HISGIIATADALGFGRKYVYAANYWQEDNTDYSAAYKLYRNDKNGKFGSIKYD
 AATSDTENSVAIYNWDEENSELHLIVKNDPDPINATFOLSKDRTYTGSRVWGFQD
 TGSIDTEAAITNNNOVTYTLPLSLAYHVLKADSTEPVNSDLVVOYKDGDRNAT
 DNOLKPHENIONKGTSPVDSLILRYFFKDSAAANGWIDWAKLGSGNIQISFGNH
 NGASDVIYELCFPSGAGIAEGGSGEIQIURMSKADWSNFEANDYSFDGAKTAYID
 WDRVTLYDQGLVWGVEP"

BASE COUNT 775 a 737 c 821 g 644 t
 ORIGIN

Query Match 14.8%; Score 602.2; DB 6; Length 2977;
 Best Local Similarity 62.9%; Pred. No. 2.5e-169;
 Matches 966; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 120 aacggcaagactattactattaaagtagatatacattcaagatcgttaagcctattagccc 179
 DB 769 ACCTCGGAGTATGTCATTTCCAGGATTAATACGAGTCGCGAAGCTGCGAGGATCAGGCC 828
 QY 180 ttatatatagcgtacaaatcaggatttgcaggcgatgaaataatggctgcagacgact 239
 DB 829 CAATATTTACGGAACCAATCAGGATCTGAGCGGACGAGAACTGGTCTATCCCGCAGGCT 888
 QY 240 tggtagcaacgaatgacggatatacaactggaacaaatgctccatgacgaagatga 299
 DB 889 CGGAGGCAACCGGCTGACGGGTATCACTGGGAGAACACATCATCGCGCGGAGGGA 948
 QY 300 ctggcagcaaatcagcgtatattatgacgaatcgttaccgcaatcgtgagcttgcaactta 419
 DB 1009 CGACAAGCGGGGGGGTGTACCGCTTTTACGATATAATCTTTGGAGATGAGAGCTTA 1068
 QY 420 ttctttagttagtgcgatgagccgggtatgtaggaagatggaacggaagatgtagca 479
 DB 1069 CTCATTGTACGCTGCAAAATGCGGGGTATGTGTCCCGGATGAAGAACGCTCCAGTTGA 1128
 QY 480 ggaagcgaagagcccttcctcgtctgtgaatcaggtcgttaacgcaacgcaaaatgcacc 539
 DB 1129 CGAGAGTGAGACGGCTCGTCACCGCGTTGGGATTAAGGTGAGTTGCCAAAATGCGCC 1188
 QY 540 gttccaaactacagcgtatgtagaatacgaatcgggtctatgtagttagttagttagt 599
 DB 1189 GTTCTCCCTTACGCTGATCTGAACCGGACGGAAGTGTATATGATGAAGAAGTTAACTT 1248
 QY 600 tttagtgacaagtagcagcactgtcaacaaagcgggggtgaaagatgtagcctcga 659
 DB 1249 CTGGTCAACCGGTATGAAACGCTTCAACGTCACCGGGATCAACGCGATTTCGTGGA 1308
 QY 660 caatgaacccgctctgtgtagcagcagccacgcatcattcattgtagaagcggagc 719
 DB 1309 TAACGAGCGCGGTGTGGTCTGAGACGATCAAGGATTCATCCGAGCAGTTTCAAGC 1368
 QY 720 gaaagagttgtagacgggtcagtcagttattccaaagctgtgaaagcattgtagcggg 779
 DB 1369 GCGAGACTCTGCTAAGACATCTGACTTGTCTCAAGCGGTGAAGAACGTCGATCCGCA 1428
 QY 780 ggcagaggttttggccgggttctttagcagattggcgccctataaagattctcaactgc 839
 DB 1429 TGCCGAATATTCGGTCTCGCCCTTTACGGTTTCGGCGCATATTTGCTCTGTCGAGACGC 1488
 QY 840 acctgattgggactctgtgaaagcaattatagctggttctgtagactattacctggatca 899

Db 1489 ACCGATTGGCCGAGTTTGGCAAGCAACTACAGCTGGTGTATTCGATTACTATCTGGATCA 1548
 QY 900 aatgcctcttagctcgcgaagtcgaagcaagagattcgtgattgattcagctacactg 959
 Db 1549 GATCAAGAAATGCTCATACGAGCAACGCAAAAGATTGCTCGATGCTGACGCTCCACTG 1608
 QY 960 gtatccccaagcgaatggcgcgagagcatcacgaattacgaat---gagtaggagcaatgacga 1016
 Db 1609 GTATCCGGAAGCACAGGCGGAGGCCAGCAATCGTCTTTTGGGGGGCGGCAATATCGA 1668
 QY 1017 aacgaagaagaacgaatgacgagcaccctcgcaccttggggagccacccatataagaaga 1076
 Db 1669 TACGCAAGAGCTCGCTACAGCGCCCAAGATCGCTATATGGGATCCGGCTTTACGAGGA 1728
 QY 1077 tagtgatcgtcctaaggaacacgagtggttttggccatactactacctcgtattgaagcagt 1136
 Db 1729 CAGCTGGATCGGCACATGGTTTTCAAGCTACTTGGCCCTTAATTCGAGAGTGGAAATCTTC 1788
 QY 1137 ggtgataataattatccggggaacaaagctggcaatgacagagtagatagctatgctgacga 1196
 Db 1789 GATTACAGAGCTATTATCCGGGTACGAAGCTGGCGATCACAGAGTTCAGCTACGCGGAGA 1848
 QY 1197 aatgatatttccggcgagattgcatgacgagtgctgctggtgatttggggcaaaatga 1256
 Db 1849 CAATCAATTTCCGGGAGGCATAGCTACCGCGAGCGCTCGGCATTTTGGAAAAATATGG 1908
 QY 1257 tgttatatggcaactactggaagctaaagatggtgtcaacaactagcttagtgcgc 1316
 Db 1909 CGTTTATGCGCGCAATTTACTTGGCAG---ACGAGGAGACATATCCGATATACAGCGCTGC 1965
 QY 1317 ttacaagcttattcgaattatgacggaaaaaactactcttctcggtgataccagtgtag 1376
 Db 1966 TTACAAGCTGTATCGCAACTACGACGCAATAAATCGGGTTTCGGCTCGATCAAGCTGGA 2025
 QY 1377 tggcaaaaatcggatattgtcaatagctggttccatgcttctgttaacgaatgacacca 1436
 Db 2026 CGCGGTACGCTCGATACGAGAACACAGCTCGGTATACGCTTCGGTAACTGACGAGAGAA 2085
 QY 1437 caaagaactgactctggttgcataaagaacgagcagcagcagcagcagcagcagcagc 1496
 Db 2086 TTCCGAAGCTCCACTGATGCTGTGATGAATAAATTTTCGAGGATCCGATCAAGCTACTTT 2145
 QY 1497 tgaatttccggcgcaagacttacatttccggtaagatgggggttgcgtataaaacag 1556
 Db 2146 CCAGCTGTCTGTGATAAAACCTACACATCCGGAGAGTATGGGGCTTCGACCAACCGG 2205
 QY 1557 ctgcgaataaagaagcagcgaacacgaatcgaatcgaatcgaatcgaatcgaatcgaat 1616
 Db 2206 ATCCGACATTTGGAACAGCAGCTATAACGAATATTAACAACAATCAATTCACGTATAC 2265
 QY 1617 cgtaccgcttggcggcagcagcagcagcagcagcagcagcagcagcagcagcagc 1651
 Db 2266 GCTTCTCCATTGTCGGCTTACCACATTTCTCTGA 2300

RESULT 2

AR012245
 LOCUS
 DEFINITION Sequence 28 from patent US 5763254.
 ACCESSION AR012245
 VERSION AR012245.1 GI:3970235
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1854)
 AUTHORS Woldike,H.Fabricius., Hagen,F., Hjort,C.Malland. and Hastrup,S.
 TITLE Enzyme capable of degrading cellulose or hemicellulose
 JOURNAL Patent: US 5763254-A 28 09-JUN-1998;
 FEATURES Location/Qualifiers
 1..1854
 /organism="unknown"
 BASE COUNT 452 a 506 c 503 g 393 t

ORIGIN

Query Match 14.8%; Score 601.4; DB 6; Length 1854;
Best Local Similarity 63.1%; Pred. No. 4e-169;
Matches 961; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 132 tattactattaaagtagatcacattcaagagctgtaagcctattagcccttatatatcagg 191
DB 81 TGTCACTTTCAGATTAAACGAGTGGAAACGTTGACGAGTACAGCCCAATATTTACGG 140
QY 192 tacaatcaggattggcgcgagcagagaaatattgctgcgagagcacttgggtgcaaccg 251
DB 141 AACCAATCAGGATCTGAGCGGAGGAGAACTGGTTCATCCGCGAGGTCGGAGCAACCG 200
QY 252 aatgaccggatatacaactggggaacaaatattgccaatgacgaagaagtactggcgaacac 311
DB 201 GCTGACGGGTTTACAACTGGGAGAACACGACATCCAGCGCGGAGAGGACTGGCTTCATTA 260
QY 312 tagcgaataactattatcagtaaatggtggcctgacacaaagcgaatgtgaaagccagg 371
DB 261 CAGCGATGATTTCTCTCGGCAACGGTGGTTCACACACACGACTGCGACAAAGCCGGG 320
QY 372 agcgtgacgactctgtttcatgacaaatcgtgaagcttggcacttattcttttagttac 431
DB 321 GCGCGTGTGTACCGCTTTTCAGATAAATCTTTGGAGAATGGAGCTTACTCCATTGTAAC 380
QY 432 gttgcgagtgccgggttatgtggttaagatggaacaggaagtgtgaggaagaaagcgaacaa 491
DB 381 GCTGCAAAATGGCGGTTTATGTCTCCGGGTAAGAAACGCGTCCAGTTGACGAGAGTGAGAC 440
QY 492 gcccctccgctcgtgtgaatcaggtcgttaaacgcgaacaaatcaccggttccaaactaca 551
DB 441 GCTCGCTCACCGGCTGGGTAAGTTCAGTTCGCAAAATGGCGGCTTCTCCCTTCA 500
QY 552 gctgatcgaatgacaaatcgggtctatgtgatgagttcgtcccaatttttttagtgaacaa 611
DB 501 GCTGATCTGAAGCAGCGACAGTGTATATGGATGAAGAAGTTAACTTCTCTGGTCAACCG 560
QY 612 gtacggcactgttcaacaaagcgggggtgaaagatatgccctcgacaaatgaacccgc 671
DB 561 GTATGAAACCGTTTCAACGTCACGGGATCAAAAGCGTATTCGCTGGATAACGAGCCGGC 620
QY 672 tctcgtgctcagacacccacgcatcattcgtgtaaaaagtcgagcgaagaagtgtgt 731
DB 621 GCTGTGCTGAGACGCATCCAAAGGATTCATCCGGAGCAGTTTACAAAGCGGAGAACTCGT 680
QY 732 agaccggtcagtcagtttatccaaagctgtgaaagcagattgacgcggggcagaggtttt 791
DB 681 CGTAAGACATCGACTTGTCAAGGCGGTGAAGAAGCTGCGATCCGATCGCGAAATATT 740
QY 792 tggcccggttcttaccgatttggcgcctataaagatatccaaactgacactgattggga 851
DB 741 CGGTCTGCGCCCTTACGGTTTCGGGCGCATATTTGTCTGCGAGGACGACCGGATTTGCC 800
QY 852 cctctgaaagcgaattatagctgttcgttagactattacctggatcaaatgcgcttag 911
DB 801 GAGTTTGAAGGCAACTACAGCTGGTTTATTCGATTACTATCTTGGATCAGATGAAGAATGC 860
QY 912 ctgcgaagtcgaagcgaagagatgctgtgatttctcagctacactggtatcccgaaagc 971
DB 861 TCATACGCAAGACGGCAAAAGATTGTCGATGCTGCGAGCTCCACTGGTATCCGGAAGC 920
QY 972 gatggcggaggcatacagaaattacgaat---gaggtaggcaatgacgaacaaagcaagc 1028
DB 921 ACAGGCGGAGGCGCAACGCAATCGTCTTGGCGGGCGGCAATATCGATACGACGAAGGC 980
QY 1029 cagaatgcagcaccctcgacaccttgggacccgacctataaggaagatagttgtagcgc 1088
DB 981 TCGGTTACAGCCCAAGATCGTATGGGATCCGGCTTACAGGAGAGACAGCTGGATCGG 1040
QY 1089 tcaatgaaacagcagagtttttggccatactactcgtcgtattgagcagcagcgtggtgataaata 1148

DB 1041 CACATGGTTTTCAAGCTACTTGGCCCTTAATTCGAGCTGCAATCTTCGATTACGACGTA 1100
QY 1149 ttatccggaacccaagctgcaatgacagtagatagctgagcgaataatgatatttc 1208
DB 1101 TTATCCGGGTACGAAGCTGGCGATACAGAGTTCAGCTACGGCGGAGACAATCACAATTC 1160
QY 1209 cggcggattgcgatgaccgatgctggtggtatccttgggcaaaaaatgatttatatgac 1268
DB 1161 GGGAGGCAATAGCTACCGCGAGCGCTCGGCATTTTGGAAAAATATGGGTTTATGCCGC 1220
QY 1269 aaactactggaactaagaagatggtgtcaacaactagtttagtgcogcttacaagtcta 1328
DB 1221 GAATTACTGGCAG---ACGGAGGACAATACCGGATTTATACGCGCTGCTTACAAGCTGTA 1277
QY 1329 tcgaattatgacgaaaaaactctacttccgtgtgataccagttgtagtgcgaacatc 1388
DB 1278 TCGCAACTACGACGGCAATAAATTCGGGTTTCGGCTCGATCAAGTGGAGCGCGCTACGTC 1337
QY 1389 ggatattgtcaatagctcgtgctccatctctgttaacgaatgcatccgcgaacaaagactgca 1448
DB 1338 CGATACGGAGAACAGCTCGGTATACGCTTCGGTAACTGACGAGGAGAATTCGGAACCTCA 1397
QY 1449 tctcgtgtcatgaataaaagcagacagcgcattcgacgcgccaatttgattcttcagg 1508
DB 1398 CCTGATCGTCTGAATAAAATTTTCGACGATCCGATCAACGCTACTTTCAGCTGTCTGG 1457
QY 1509 cgcgaagacttacttcccggttaagttggtggttcgataaaacagctcgcgaataaa 1568
DB 1458 TGATAAAACCTACATCCCGGAGAGATATGGGCTTCGACCAAAACGGATCCGACATTA 1517
QY 1569 agaagcagcgcacacacgcaaatcttcaggcaaccgttttacttataccogtaccgcctt 1628
DB 1518 GGAACAAGCAGCTATACGAATATTAACAACAATCAATTCACGTATACGCTTCTCCATT 1577
QY 1629 gacgcaatatacaattgtgctga 1651
DB 1578 GTCGGCTTACCACATTTGTTCTGA 1600

RESULT 3

LOCUS I14943
DEFINITION Sequence 28 from patent US 5457046.
ACCESSION I14943
VERSION I14943.1 GI:1249851
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Woldike,H.F., Hagen,F., Hjort,C. and Hastrup,S.
TITLE Enzyme capable of degrading cellulose or hemicellulose
JOURNAL Patent: US 5457046-A 28 10-OCT-1995;
FEATURES Location/Qualifiers
source
1. .1854
BASE COUNT 452 a 506 c 393 g
ORIGIN

Query Match 14.8%; Score 601.4; DB 6; Length 1854;
Best Local Similarity 63.1%; Pred. No. 4e-169;
Matches 961; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 132 tattactattaaagtagatcacattcaagagctgtaagcctattagcccttatatatcagg 191
DB 81 TGTCACTTTCAGATTAAACGAGTGGAAACGTTGACGAGTACAGCCCAATATTTACGG 140
QY 192 tacaatcaggattggcgcgagcagagaaatattgctgcgagagcacttgggtgcaaccg 251
DB 141 AACCAATCAGGATCTGAGCGGAGGAGAACTGGTTCATCCGAGGCTCGGAGCAACCG 200
QY 252 aatgaccggatatacaactggggaacaaatattgccaatgacgaagaagtactggcgaacac 311

Db 201 GCTGACGGGTACAACTGGGAGAACACGATCCAGCGCGGAAGGACTGGCTTCATTA 260
QY 312 tagcgaataatttttagtaagtaattggtgacctgacacaaagccgaatgtgaaagccagg 371
Db 261 CAGCGATGATTTCTCTCGCGCAACGGTGGTGTTCAGACACCGACTCGGCAACGCCGG 320
QY 372 agcggtagacacttcgttcacgaacacgcgtgaagcttgacattatttttagttac 431
Db 321 GCGGTTGTATACCGTCTTTCAGGATAAATCTTTGGAGATGAGGCTTACTCCATTGTAAC 380
QY 432 gttccgatggcgggttatgtgctaaagatgaaacggaatgtgcaggaagcgaataa 491
Db 381 GCTGCAATGGGGGTATGTCTCCGGGATGAAGACGGTCCAGTGTGACGAGAGTGAGAC 440
QY 492 gggcccttcgctgttgggaatcaggtcgttaaacgccccaaaaatgcaccgtttccaaactaca 551
Db 441 GGCTCGGTCAACCGCTTGGGATAAGGTGAGTTTGCACCAAAATGCGCGCTTCCCTTCA 500
QY 552 gctgatcgaatgacaatcgggtctatgtgatgagttcgtccatttttttagtgaaca 611
Db 501 GCTGATCTGAACGACGGCAAGTATATATGATGAAGAAAGTTAACTTCCCTGGTCAACCG 560
QY 612 gtacggcactgcttcaacaaaggcggtgaaagatatgccctcgaacaatgaaccgcg 671
Db 561 GTATGGAAACGCTTCAACGCTCAACGGGATCAAAACGGTATTCGCTGGATAACGAGCGCGC 620
QY 672 tctctggtcgatcagcaccacgcattcattggtgaaagtcgagcgaaagagttggt 731
Db 621 GCTGTGCTGAGAGCGCATCCAAAGGATTATCCGGAGCAGTTACAAAGCGGCAAGACTCGT 680
QY 732 agaccggtcagtcagtttccaaagctgaaagcgtgaaagcgtgacgcggggcagaggtttt 791
Db 681 CGCTAAGACATCGACTTGTCAAGGCGGTGAAGAACGTCGATCCGATCGCGCAAAATATT 740
QY 792 tggcccggttcttaccgatttggccctataaagatttcaaatgcacctgattggga 851
Db 741 CGGTCCTGCCCTTACGGTTTCGGGCGATATTGTCCTGAGACGACCGCGGATGGCC 800
QY 852 ctctgtaaaggcaatttagctggttctgtagctatttaccctgatacaaatgcgccttag 911
Db 801 GAGTTTGAAGGCAACTACAGCTGGTTTATCGATTACTATCTGGATCAGATGAAGAATGC 860
QY 912 ctccgaagtgcgaagcaagatgctggtgatttgcagctacacagtgatcccgaggc 971
Db 861 TCATACGACGACGACGCAAAAGATGCTCGATGCTGGACGTCACCTGGTATCCGGAAGC 920
QY 972 gatggcgaggatcagcaattacgaat--gaggtaggcaatgacgaacgaagaagc 1028
Db 921 ACAGGGCGGAGCCAGCGAATCGTCTTTGGGGGGCGGCAATATCGATACGACGAGAGC 980
QY 1029 cagaatgcaggaactgcacctgtgggaccgcacataaaggaagatggtggatcgc 1088
Db 981 TCGCGTACAAAGCGCAAGATCGCTATGGGATCCGGCTTACAGGAAGACAGCTGGATCGG 1040
QY 1089 tcaatggcaacgcaggttttgcctactactacctcgattgaagcagtcggtggaataa 1148
Db 1041 CACATGGTTTCAAGCTACTTGCCTTAAATTCGAAGCTCAATCTTCGATTCAGACGTA 1100
QY 1149 ttatccgggaacaaagctggaatgacagagtagtatggtggcggaataatgatatttc 1208
Db 1101 TTATCCGGGTACGAAGCTGGCGATACAGAGTTTCAGCTACGCGGAGACAATCACATTTC 1160
QY 1209 cggcggtatcgatgaccatgctggttatctctggcattctggcgaataatgatattatgc 1268
Db 1161 GGGAGGCATACCTACCGGGACGCGCTCGGCATTTTGGAAATATGCGCTTATGCCGC 1220
QY 1269 aaactatggaagctaaagatggtgtcaacaactacgttagtgccgcttacaagctta 1328
Db 1221 GAATTAAGTGGCAG--ACGGAGGACAATACCGATTATACCGAGCGCTCTTACAGCTGTA 1277
QY 1329 tcgcaattagcggaaacaaacttacttctcgtgtatcaccagttgttagtgcgaacatc 1388

Db 1278 TCGCAACTACGACGCGCAATAAATCGGGTTCGGCTCGCATCAAAAGTGACGCGCTACGTC 1337
QY 1389 ggaattgtcaatagctcggtccatgctctctgttaacgaatgcataccgacaaagactgca 1448
Db 1338 CGATACGGAAGAACAGCTCGGTATACGCTTTCGATTAACGTACGAGGAGAAATCCCGAACCTCCA 1397
QY 1449 tctggtgtcatgaataaaagcatggacagcgcatcttcagcgcccaatttgcattcttcgg 1508
Db 1398 CTTGATCGTGTGATGAATAAATTTTCGAGGATCCGATCAACGCTACTTTCAGCTGTCTGG 1457
QY 1509 cgcgaagacttacattcccgtaaaagtatgggggttcgataaaacacagctcgcgaataa 1568
Db 1458 TGATAAAACCTTACACATCCGGGAGATATGGGCTTCGACCAAAACCGGATCCGACATTC 1517
QY 1569 agaagcagcgcccaatcacgaatatttcagcaacccgttttacttataccgtacgccttt 1628
Db 1518 GGAACAAGCAGCTATTAACGAATATTAAACAACAATCAATTCACGTATACGCTTCTCCATTT 1577
QY 1629 gacggcatatcacattgtgctga 1651
Db 1578 GTCGGCTTACCACATTTGTTCTGA 1600

RESULT 4
LOCUS 173663
DEFINITION Sequence 28 from patent US 568593.
ACCESSION 173663
VERSION 173663.1 GI:3009804
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Woldike,H.Fabrizius., Hagen,F., Hjort,C.Malland. and Hastrup,S.
TITLE Enzyme capable of degrading cellulose or hemicellulose
JOURNAL Patent: US 568593-A 28 11-NOV-1997;
FEATURES Location/Qualifiers
source 1..1854
BASE COUNT 452 a 506 c 393 g
ORIGIN

Query Match 14.8%; Score 601.4; DB 6; Length 1854;
Best Local Similarity 63.1%; Pred. No. 4e-169;
Matches 961; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 132 tattactattaaagtagatatactcaagatcgtaagcctatagcccttatatacgg 191
Db 81 TGTCACTTTTACGATTAATACGACGTCGGAAGCTGCAGCGATCAGCCCAATATTTACGG 140
QY 192 tacaatcaggatttggcagggcgatgaaatatggctgcacagcagcttggtggaacgc 251
Db 141 AACCAATCAGGATCTCAGCGGACGAGAACTTGGTATCCCGAGGCTCGGAGGCAACCG 200
QY 252 aatgcagcgatatacaactgggaacaaataatgtccaatgcaggaagtgcactggcgagcaatc 311
Db 201 GCTGACGGGTTTCACTGGGAGAACAAACCATCCAGCGCGGAAAGGAGCTTGCTTCATTA 260
QY 312 tagcgaataactttatgcagtaattggtgctgcacaaagccgaatgtgaaagccagg 371
Db 261 CAGCGATGATTTTCTCTCGGCAACGGTGTTCAGACACCGACTCGGCAACGCCGG 320
QY 372 agcggtagcacttcgtttcattgaccaatcgctgaagcttggaacttattcttagttac 431
Db 321 GCGGTTGTTTACCGCTTTTTCAGATAAATCTTTGGAGATGAGAGCTTACTCCATTGTAAAC 380
QY 432 gttgcgagtagcggttatgttggttaagtaggaaacggaagtgtgcaggaagcgaataa 491
Db 381 GCTGCAAAATGGCGGGTTATGTGTCCCGGATGAAGAACGCTCCAGCTTGACGAGAGTGAGAC 440
QY 492 gggcccttcgctcgttgggaatcaggtcgttaaacgccccaaaaatgcaccgtttccaaactaca 551

Db 441 GGCCTCCGTCACCGCGTGGGATAAGTTCGAGTTTGGCCAAATAATGCGCGTCTCCCTTCA 500
Qy 552 gctgatcgaatgaacatcggtctctatgtgtagtgagtcgtccatttttagtgaacaa 611
Db 501 GCCTGATCTAACGACGACAGTGTATATGATGATGAAGAAGTTAACTTCCTGCTCAACCG 560
Qy 612 gtacggcactgtccaacaaagcggggtgaaagatatgcccctgcacaaatgaaccgcg 671
Db 561 GTATGGAACGCTTCAACGCTCAACGGGATCAAGCGATCAAGCGTATTCGCTGGATACGAGCGCGC 620
Qy 672 tctctggtcgatcgacccaccacattcattggtgaaaaagtcgagcgaaagattggt 731
Db 621 ECTGTGCTCAGACGATCAAGGATTCATCCGAGCAGTTACAGCGCGCAGAACTCGT 680
Qy 732 agacgggtcagtcagtttatccaaagctgtgaaagcgattgacgcggggcgagagtttt 791
Db 681 CGCTAAGAGCAGTACACTTGTCAAGGGCGGTGAAGAAGCTCGATCCGATGCCGAATATT 740
Qy 792 tggccggtctttacgagattggcgctataaagatctcaaaactgcactgaattgga 851
Db 741 CGTCTCCCTTACCGGTTTCGGCGCATATTTCTCTGAGGACGACCGGATTCGCC 800
Qy 852 ctctgtaaaagcaattatagctgtgttagactattacctggtatcaaatgcgccttag 911
Db 801 GAGTTTGAAGCAACTACACTGCTGTTATCGATTACTATCTGATCAGATGAAGATGC 860
Qy 912 ctctgcaagtcgaagcaagattgctggtatgtattgacgacgacggtatcccgagc 971
Db 861 TCATACGAGCAACGCAAAAGATTGCTCGATGCTGTCGACGCTGCTGATCCGGAAGC 920
Qy 972 gatggcgagcagcacaataatgaat--gagtgagcaatgcaaaacgaagaagc 1028
Db 921 ACAGGGCGGAGCCAGCAATCGCTTTGGCGGGCGGCGCAATATCGATACGCAAGGC 980
Qy 1029 cagaatcgagcagcactgctggtggacccgacctataagaagaatagttgagtcgc 1088
Db 981 TCGCGTACAAGCGCCAAAGATCGCTATGGATCCGCGTTTACCAGCAAGACAGCTGATCGG 1040
Qy 1089 tcaatggaacagcgaggttttggccactactacctcgtatggaagcagtcggtggaataa 1148
Db 1041 CACATGGTTTTCAAGCTACTTGGCCCTTAATTCGGAAGTGAATCTTCGATTCAGAGCTA 1100
Qy 1149 ttatccgggaacaaagctggaatgacgagatagctatgctgagcgcaaaatgatattc 1208
Db 1101 TTATCCGGGTACGAAGCTGGGATACAGAGTTTACGCTTACGCGGAGCAACATCATCTTC 1160
Qy 1209 cggcggtgattgcatgacgagtgctggtggtatcttggcgcaaaatgatgttttatggc 1268
Db 1161 GGGAGGCATAGCTACCGCGGACGCGCTCGGCATTTTGGAAATATGCGGTTTATGCCGC 1220
Qy 1269 aaactactggaagctaaaggatggtgtcaacaaactacgttagtcgagcgttacaagcttta 1328
Db 1221 GAATTTACTGGCAG---ACGGAGGACAAATACCGATTTATACCGAGCTGCTTCAAGCTGTA 1277
Qy 1329 tcgcaattatgacggaataaaactactcacttcggtgataccagtgtagtcgcaaaacac 1388
Db 1278 TCGCAACTACGACGCGCATTAATCGGGGTTTCGGCTCGATCAAGTGGCGCGCTACGTC 1337
Qy 1389 ggatattgtcaatagctcgttcattgcttctgtacgaatgcatccgcaaaagaaactgca 1448
Db 1338 CGATACGGGAACAGCTCGGTATACGCTTCGTTAACTCAGCAGGAGAAATCCGAATCCA 1397
Qy 1449 tctcgtgtcatgaataaaagcatgagcagcgattcgagcgcgaattgatttcttcgg 1508
Db 1398 CTGTATCTGCTGTAATAAAATTTTCGACGATCCCGATCAACGCTACTTTCCAGCTGCTCGG 1457
Qy 1509 cgcgaagacttaccattccggttaaagtagtggggttcgataaaacagactcgcaaatata 1568
Db 1458 TGATAAACCTACATCTCGGAGAGATATGGGCTTTCGACCAACCGGATCCGACATPAC 1517
Qy 1569 agaagcagcgcaatcacgcaaatattcaggcaacggttttactataccgtaccgctttt 1628

Db 1518 GGAACAAGCAGCTATTAACGAATATTAAACAATCAATTAACGATACGCTTCTCCCTCA 1577
Qy 1629 gacggcatatcacattgtgctga 1651
Db 1578 GTCGGCTTACCACATTGTTCTGA 1600

RESULT 5
BACCELAA
LOCUS BACCELAA
DEFINITION Bacillus lautus endo-beta-1,4-glucanase (cela) gene, complete cds.
ACCESSION M76588
VERSION M76588.1
KEYWORDS cela gene; endo-beta-1,4-glucanase.
SOURCE Bacillus lautus (strain PL236) DNA.
ORGANISM Paenibacillus lautus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Paenibacillus.
REFERENCE 1 (bases 1 to 2831)
AUTHORS Hansen, C.K., Diderichsen, B. and Jorgensen, P.L.
TITLE cela from Bacillus lautus PL236 encodes a novel cellulose-binding
endo-beta-1,4-glucanase
J. Bacteriol. 174, 3522-3531 (1992)
JOURNAL
MEDLINE
FEATURES
Location/Qualifiers
1..2831
/organism="Paenibacillus lautus"
/strain="PL236"
/db_xref="taxon:1401"
<1..438
/note="ORF"
/codon_start=1
/transl_table=11
/protein_id="AA22302.1"
/db_xref="GI:142662"
/translation="LQYRAADTNAADNQTSPFNKNNKTSVADLSTLKIRYFTKDG
SAAVNGWIDWAOLGGSNIQISFGNHTGTNSDTYVELSPSSSAGSIAAGQSGETOLRM
SKTDWSENEANDYSFDGTTKTAFAWDWRVLYQNGOIVMGTA"
518..522
/gene="cela"
518..2633
/gene="cela"
531..629
/gene="cela"
531..2633
/gene="cela"
/EC_number="3.2.1.4"
/codon_start=1
/transl_table=11
/product="endo-beta-1,4-glucanase"
/protein_id="AA22303.1"
/db_xref="GI:142663"
/translation="MKTRORKLFLVSAALAVSLTWTVPMPASVNAASDVTTFTINTQS
ERAAISPIYGTNQLDSTGNTENNRRGRLTGYNNENASSAGRDMLHYSDDFLCG
NGVDPDQCDKPCAVVATFPHDKSLENGAYSVTLQWAGYVSRDKNGPYDESETPSR
WQKVEFAKNPESLOPHLNDGVYMDVEVNLVNRVYGNASTGTGKAYSLDNEPALWS
ETHPRHPELOAAELVAKSIDLSKAVKNVDPAETFGFALYFGAYLSLQDAPGWPS
LGNTYSEFDYDLDOMKNAHTQNGKLLDLVHMVPEAQGGQRIVFGAGNIDTOK
ARVQPSRLWDFATQDSWIGTWSYSLPLIKLQSSITOTYPTGTLKAITESYGGDN
HISGGIATADALIFGKYGVYAANYWQEDNTDYTSAAKLYRNYDGNKSGFSLKVD
AATSDTENSALVYSYDEENSELHVLNKNFDDPINATFQLSGDKTYTSGRWGFDQ
TGSDI TEQAATINNNQTYTLPLPSAYHVLKADSTEPVNSDLVVOYKDGRRNAT
DNQIKPHFNIONKTSYDLSLTLYFTTKDSSAMNGWIDWAKLGSNIQISFGNH
NGADSTYAEELGFSGAGSIAEGGSGEIQLRMSKADWSNENFANDYSFDGAKTAYID
WDRVLYQDGLVWGTPEP"
630..2630
/gene="cela"
/product="endo-beta-1,4-glucanase"
BASE COUNT 742 a 700 c 784 g 605 t
ORIGIN
Query Match 14.8%; Score 599; DB 1; Length 2831;


```

QY 1464 taaagcatgacagcgcatcgagcccaatttgatatttcccggcgcgaagacttacat 1523
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4254 CAAGAACTATGACAGCGACGACATTCATTTCTCAATTGACAGCAGCAAGAACTACAC 4313
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1524 ttccggttaagtgtgggttcgataaaacagctcgcaaatataaagaagcagcgccaat 1583
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4314 AATAGGAATGTATGGGATTTGACAGAGGAGCTCCCAATATTTACTTAAAGAACTCCAT 4373
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1584 cagcaaatcagcaaccgtttacttataccgtaccgctcttgacggcatcacat 1643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4374 ACTGAACATAAAGACAATACCTTCACATATACAGTACCGCGCTTGACAGCGTCCATAT 4433
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1644 ttgtct 1649
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4434 TGTGCT 4439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
AE007607 13603 bp DNA linear BCT 27-JUL-2001
LOCUS Clostridium acetobutylicum ATCC824 section 95 of 356 of the
DEFINITION complete genome.
ACCESSION AE007607 AE001437
VERSION AE007607.1 GI:15023810
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum
          Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
          Clostridium.
REFERENCE 1 (bases 1 to 13603)
AUTHORS Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
          Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
          Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
          Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
          Genome Sequence and Comparative Analysis of the Solvent-Producing
          Bacterium Clostridium acetobutylicum
          J. Bacteriol. 183 (16), 4823-4838 (2001)
          21359325
          11466286
REFERENCE 2 (bases 1 to 13603)
AUTHORS Childress,D., Zeng,Q. and Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production,
MEDLINE Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
          Beaver Street, Walcham, MA 02453-8443, USA
FEATURES
          source
            1. .13603
               /organism="Clostridium acetobutylicum"
               /strain="ATCC 824"
               /db_xref="ATCC:824"
               /db_xref="taxon:1488"
               75. .2255
               /gene="CAC0911"
               75. .2255
               /gene="CAC0911"
               /codon_start=1
               /transl_table=11
               /product="possible processive endoglucanase family 48,
               secreted; Celf ortholog; dockerin domain"
               /protein_id="AAK78887.1"
               /db_xref="GI:15023811"
               /translation="MLKISKPKKINAVALTSTVLFGLSLGILLTNKVAATTTDSSLK
               VDNAYTQRETHYKMKHDANGYFSGDGVYPSVETFWAEADYGHETSEAFSIYMW
               LEAMQKRTIGSGVNTAWDAETKYMIPSHQDQPMGRYNASSPATYSPEDWPSKYP
               SRWDQGAAGQDPISEDLKSAVGTSDMYGMHMLDVMYNGFNHEDGTSKNYINTY
               QREGEVETVPPQCDWAFKYGGKNGYLDFTGNSYAKQAKYTDAPADARAQAT
               YEAAQAKEDVDLSIVGKASKMGYLRVAMFDKYFRKIGNSTOAGNGKDSMWKSLTT
               WYVWAGSSNNDSWKIGCSHSHFGYONPLTAWLSTDSQFKPKSATCATDWAASLT
               QVDYQRLQSSAAGIAGSSNHGRIEAWPEGTATFDGNGQEPEVFPKDSNTWFG
               MQSWQRQAQYITQSKDKAKALLDKWVKIKSVVKNPNPAGTFFVPSKLSWTGP
               TWGTSGTGNPNLHVNDVSTYTDIGTSSSTADALAYAAATGDKDQSLSKTILDDIW
               KNYQDAKGSAPQMDSYRVFQEVYIPQGTGTWPNGDVYIKSGNKFIDIRSQYKNDP
               DYARVKSDEVAGKSSFTNHFREWAESYAIANANYGTLFANTATPGDVNGDGVVNGRD

```

gene

CDS

LMELROYLAGKIDASKINIAAADVNNNDGVVNGRDIMELTKLIAK"

2337. .3785

/gene="CAC0912"

2337. .3785

/gene="CAC0912"

/codon_start=1

/transl_table=11

/product="Possible non-processive endoglucanase family 5,
secreted; Celf homolog secreted; dockerin domain"

/protein_id="AAK78888.1"

/db_xref="GI:15023812"

/translation="MAACLAFCNVAAYADDVNVNSNDYLHSDGSKLLDDYGNQVPMVTG
IAWGLETPNYCFHGLWANRDLNINLIVADNGFNTLRPLSVELVNMQRQGYVPTPDS
INDYISPELKGNSLIQILDVIAYSKYGKVKVMLDMHRIESGGOTATWTYKTYTDDY
EKCWOYLADRYKNDTIVIAADIFNEPHGKAYRAETSAKNDTDDNNRYEAEKVGKK
ILDINPKMLIVVEGVETYPKEGTAAGSTNPDDYGGWGGNLRGVKDYVDLAPVKNQ
VYSPHDYGPVGSVDQTFDQGTFTQESLLNDINRPSWFIQEKNTIAPLLIGWGMNDG
GRNEQMTDMAKLSIDNKNHFTWCLNANSQDGTGILEYDFKTDIKKLALVQPTLWK
DDASGKFELDHENGLKNGTHVQPTTIKVKQGDVNGDGVNGRDMELRKYISQGI
TTINKDVAIDNGDGVNGRDLVELIKMISNQ"

3906. .6044

/gene="CAC0913"

3906. .6044

/gene="CAC0913"

/codon_start=1

/transl_table=11

/product="Possible non-processive endoglucanase family 9,
secreted; Celf ortholog; dockerin and cellulose-binding
domain"

/protein_id="AAK78889.1"

/db_xref="GI:15023813"

/translation="MVKRLKSLIAVFITTMAGMGFMNLTGDKVMADETNTNFYGEA
LOKSLMFYEFERSGKLPODKRONWRGDSGLDDGVDGLDTGGWYDAGDHVKENLPAS
YTLMLGWSLVQRNDSYKSGOLKYLTDKMKWCDFLMKCHPAPDVEYVOYVGDGLDH
KWWGPAEYMQDRPSFKYDDNDPGSAVTAESAALATALFNKDDADYANKCLSHAK
DLNFADKTRSDKGYTAANGYISSTSFYDLSLWACWLYMATNDRSYLDKAEAYVNNV
SREOQTDIISYKMGCDWDVHYGAQVLLAEITNKQIYKDSVERNLVDYTVGYDGNKVQ
YTPKGLAWISSWPLRYLATAFLADTYSKWSGGDASKAKAYEDFAKSDQDYALGSSG
RSYVGVGVNPPHPHRTAESWFDKTVPGYSHRHTLYGAMVGPDPQNDKFDVVAN
FQNEPACDYNAGLAVTLSSMGYGGKPIDNFNAIEKPTDEFVVEAAVATKNFE
ETKSIYNTKGPARTDKLSFKYIDISEFVKOGYKASDIKVTYNEGTSQILIP
WDEKNIIYVNVDFGTGLYFGGGSAYKKEIQFMTAPDNATWNSDNFSYTGVEETP
GDTPLDKNIPYDAGVKVYGNPEKNVVVVDINGDGVNGRDLMLRKYISQITDI
NKDSADINGDGVNGRDLMLRKYISVNP"

6080. .6625

/gene="CAC0914"

6080. .6625

/gene="CAC0914"

/codon_start=1

/transl_table=11

/product="Cellulosome integrating cohesin-containing
protein, secreted"

/protein_id="AAK78890.1"

/db_xref="GI:15023814"

/translation="MKKGIITAVILSLVITGVCKSNDRKTSVSNLVSSSEKSNLK
ISSSGKTSNVEVKVKNVAKNVCICDPKIKYDASKLDVSGISPEILKDPKDL
EYNVDKNGVITILYSYSDKNIGRELISKNGDFVTLNLIKDDAKKGTQISFNGNP
FYDNKQKGVSYTTNKGEIEIK"

6638. .8458

/gene="CAC0915"

6638. .8458

/gene="CAC0915"

/codon_start=1

/transl_table=11

/product="Endoglucanase A precursor
(endo-1,4-beta-glucanase) (cellulase A), secreted;

/protein_id="AAK78891.1"

/db_xref="GI:15023815"

/translation="MKRNKILVLSALATAFSSMLVTPYSGLVKFAFDADVNVNIDTNAE
KQATSPYTGNDQFSNAKVTAIGGNRSTGYNNWNNDSNAGTDMKNESNLTLY
DVPEKYNPAVSVYTAHDKSLAMGVPSYLVLTQAGGVAAQDQSGPLANTDVPSSKW
KKVEFNKGPLSLPTDITDGSVYMDGFVNYLVNKGYSASGSKGIGKISLDNEPSLWFS

```

THPLIHPDKTSEVLDDQLAQVKKIDPAEFTFGPALFSGFAENRPNFNSPDWSSV
KGYOWEIDYLDNKKNSDAACKRWGLDLDLHWYPEAKGGQORVTSRDNVDCNKA
RQAPRLMSWSTEDSWGCKRWGLPLPKVKSISDKTYPGTKLKSYSINTGGEDH
ISGDIADGALGVFGKYVGFATYECNSDKNNYVSFAFLNNYDGNNSKYGDYVK
CSDSDINSSDTSVTSNDGNKMDIIIVNNKYTDSINFNFNSSNNKYTSQVWGFDS
NNSNITRDSVSSISGNKFTYKIPALFAVHIVLTAAQSSVKGDVNGDVVNGRDLME
LRKYLAGSTNSIDLNAADLNDGAVNGRDLLELIKLVFK"
8489..10567
/gene="CAC0916"
8489..10567
/gene="CAC0916"
/gene="CAC0916"
/codon_start=1
/transl_table=1
/product="Possible non-processive endoglucanase family 9,
secreted; Celg ortholog; dockerin and cellulose-binding
domain"
/protein_id="AAK78892.1"
/db_xref="GI:15023816"
/translation="MKLLATLVTFVFMICMQLTPVQAAKSTEDKNYNYVDAFSKILF
YEANWCGPDAGNNLRKWRGCHEDDGDVGLDLTGGPHDAGDHVKFGLPOAYASTLG
WAYVEEDSFVKKGDKYMLNLIKHFDTYFLKCYPNKTFYQCGDGTTHSHWGPPE
LQTLDIRPLLYAATPDKPSDVCGSTAAALALMYINYKDVDAEYANKCLTAAKDLF
DFGTYIGLSSEGGYFGSTGYLDDLWSGAVMLYAAATNDNSYMTDIDSLFAKGGDN
SYSNHWTCHWDVDFGVFKLAQLTNPKYKATAEENLNTWKIDAPTGGGLYIASW
GTLRYTAECLMALVYKTSKNEEYLNAFASQIDYMLGSNPNRSYEVGFGNVPKNP
HRAASGRMECAPYVEKKTDEPKHLLYALVGPDPGTSDDYSDVNOYVNSEVADYN
AGFVAGLMSAGNYGGDNLDPEDVPGIEANSPOYYVDANIMKSENEECTIOAFINASL
LPPHYKMSRFRYVDLSLYKGGTYVNDVGVRYAAGGTLISGFIPYDEAKHIYV
EASMPGQYGAEMQFAITAYNAKCLDASDNFSRNLKSDSTNISENIPYKNGKIF
GNFSGVEVTKGVDNGDNTVDGRDLMELRYLAGNNPNIINLKAADLNGDGVINGRDLV
ILIKLSGQ"
10624..12237
/gene="CAC0917"
10624..12237
/gene="CAC0917"
/codon_start=1
/transl_table=1
/product="and cellulose-binding endoglucanase family 9;
Cell ortholog; dockerin domain"
/protein_id="AAK78893.1"
/db_xref="GI:15023817"
/translation="MKLLIACLTVMLLIGTQIKPLETITADTNDKHNKYVDAFSKIMF
YEANWCGADAGNNLRKWRSGCHENDGDVGLDLTGGPHDAGDHVKFGLPOAYASTLG
WAYVEEDSFVKKGDKYMLNLIKHFDTYFLKCYPDNNTFYQCGDGTTHSHWGPPE
LQTSRPLTYKATPSTPASDVCGNTAAALALMYINYKDVSAVANKCLTAAENLYNLG
KNYKMSQSGFYSSGYLDDLWSGALWLYEATNNNLYSDAQSFLOSANINEYQYN
WTHCWDDVILGVGVFKMAQITGDARYKNIVGNIDYTNVSTVTPGGLKFRFTGWTLYR
TAAECMALVYKTSNDSEANLAKSQIDYILGTNPQMSYEVGFGSKPKYPHHRA
SGRNESAGEKTEPEKHLILYALVGGPDQDNTYENIEEYOHSEVADYNAGFVGALA
GISNYFGONQSVKINDPDVKDSNTTNTILGLDNDGVVNGRDIIVMMRQYLAGKTVSG
IDKNALDINGGAVNGRDLMEIKKVSNN"
12255..13532
/gene="CAC0918"
12255..13532
/gene="CAC0918"
/codon_start=1
/transl_table=1
/product="Possible non-processive endoglucanase family 5,
ortholog of mannase A, secreted; dockerin domain"
/protein_id="AAK78894.1"
/db_xref="GI:15023818"
/translation="MKLLKIVSTVAVGVLSTLVSFSSITKVKAAADTDNKPGINMWN
GSKYFPLGANYAWDEWDDNDNDWNTTRPAKIKADFDNMSAQGHVTRVWVFCNNYAS
PLFSSQDGKVCYGLPDKWTDHMEADYAYSKNMKIYFTLTSDFDVAKTNNSPFVHGS
IDDPTRKSYIDNAVTPVVKALDGNPVMGWDVINEPEWTISSADGNGPDGSKWGL
STLRSFVVDVDCIHOYAKQPVSVGSLSKLKMGQYDFWGLDLYDFHWDWATPY
FNPLKTPYSQKAFKPKVPIIGEMNPDTONSLSKMSHKOVLDGLVNGYSGYMLAWT
DASVNCVKTPADPEKTEPELPIDMPMPGLKGDVNGDGVINGRDLMLVRLQYLAG
QSVNLKANTDVGNGDVVNGRDLMEIKVIMSK"
4869 a 1875 c 2928 g 3931 t

```

BASE COUNT
ORIGIN

```

Query Match 11.2%; Score 455.8; DB 1; Length 13603;
Best Local Similarity 57.4%; Pred. No. 3.9e-125;
Matches 883; Conservative 0; Mismatches 647; Indels 9; Gaps 3;

QY 125 caaagactattactattaaagtagatatacgaagatcgtaagcctatttagccctata 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6732 CAGCGGATGTAATATGTAATATGTAACAAATGCTGAAAAAAGCTATTAAGTCCATATA 6791

QY 185 tatacggtaacaaatcaggatttggcaggcgagataaataatggctgcagacgacttgggtg 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6792 TATATGGTACAAATAGGATTTTA--GCAATGCTAAGGTACAGCAAGAAAGAAATTTGGTG 6848

QY 245 gcaacgaatgacgcgatacaactgggaaacaatatgtccaatgcaggaagtgaactggc 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6849 GAAATAGATCTACAGGATACAAATGGGAAAAATATGATTCGAATGAGGAAACAGATTGGA 6908

QY 305 agcaatctagcataactattatgcagtaattgtggtgcctgcacaaagccgaatgtgaaa 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6909 AGAATGAAGCGATAATATTATGGCTTACACTTTATGATGTTCCTAAGAAAAAGTATAATG 6968

QY 365 agccaggagcggtagcgaacttcttcaatgaccaaactgcgt--gaacttggcacttatt 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6969 AGCCAGCTTCAGTTTATACAGCTTTCCATGATAGTCACCTTGCATGGGAGTACCATACT 7028

QY 422 ctttagttacgttgcgcgattggtccgttatgtggttaaggatggaacggaagtgtgcagg 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7029 CACTTGTAACTTCAAGCAGGTGGATATCTTCGAGCTGATCAGAGTGGTCTCTTTGCCAA 7088

QY 482 aaagcgaagagcccttccgctcgttggtaactcaggtcgttaaacgccaacaaatgcacct 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7089 ATACAGATGTGGCACCATCATCAAAATGAAAAAGTTGAGTTTAATAAGAAATGGCCCAT 7148

QY 542 tcaactacagcctgactgaatgacaaatcgggtcgtatgtggatgaacttcgtccatttt 601
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7149 TATCACTTACCGAGATACAAACAGATGGCTCTGTTTACATGGAGTGGTGTGAATATC 7208

QY 602 tagtgaacaagtacgcgcactgtcttcaacaagcgcggggtgaaaggtatgccctgcaca 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7209 TTGTAATAATATGGAAGTGCATCAGGTTTCAAGAAAGAAATTAAGGGATATTCATTAGATA 7268

QY 662 atgaacccgcctctctgtgctgcatacgcacccacgcattcattgtgaaaaagtcggagcga 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7269 ACAGAGCTTCACTTTGGCCAAAGCACACATCCACTTATTCATCCAGATAAAACTAAATGTT 7328

QY 722 aagagttggttagaccgttcagtcagttatccaaagcgtgaaagcgtgaaagcgtgaaagc 781
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7329 CTGAAGTATTGGAATAAGGATACGACGCTTGCACAAAGTTGTAAGAAATAAGATCCACAG 7388

QY 782 cagaggtttttggcccggttctttcacgatttggcgccctataaagatcttcaaacgtcac 841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7389 CTGAACCTTTGGACACGACATTTGTTGGTTTATGTCGATTTACGATTTCAATAGTTTAC 7448

QY 842 ctgattgggactctgttaaaaggcaattatagctggttcgtagactattacctggatcaaaa 901
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7449 CTGACTGGACGAGTGTAAAGGAAATTTATCAGTGGTTTATAGATTACTATCTTCTGATAATA 7508

QY 902 tgcctcttagctcgaagtcgaagcgaagattgtcgtgattgttcgacgtacactcgtt 961
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7509 TGAAGAAAACTAGATGCTGCAGGAAAAAGACATTTTAGATGTCATTTGGATTTCTTCTTGGT 7568

QY 962 atcccgaagcgtggtggcgaggaatacgaattacgaa---tgaggtgaggaatacgaagaa 1018
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7569 ATCCAGAGGCTAAGGTGGCGGACAAAGAGTAACAACATAGTAGTACTTCTTAATGTAGATT 7628

QY 1019 cgaagaagccagaatgcaggcacctgcacaccttgcctgttggtggcccgccacctataaagaata 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7629 GTAATAAAGCAAGAATGCAGGCTCCAAGAGTGTATGGGATTTCAACTTACACAGAGACA 7688

QY 1079 gttggatcgtcgaatgaaacagcagatttttggccatctactacctcgtattgaagcagtcgg 1138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7689 GCTGGATAGGTGCTAGTGGTGAAGTGGGAGTTGGCCACTTAATACCTTAAGGTTTAAATCATCAA 7748

QY 1139 tggataaatattcccggggaaccaagctggcgaatgaccgagtagctatgctatggcgcgaaa 1198

```


JOURNAL Submitted (31-OCT-1995) Moreland D. Gibbs School of Bio Sci,
 Macquarie University, Sydney, NSW 2109, Australia
 FEATURES Location/Qualifiers
 source 1. .3900
 /organism="Caldicellulosiruptor saccharolyticus"
 /strain="Rt8B.4"
 /db_xref="taxon:44001"
 /note="thermophilic anaerobe"
 819. .3554
 /gene="manA"
 819. .3554
 /gene="manA"
 /EC_number="3.2.1.78"
 /codon_start=1
 /transl_table=11
 /product="beta-1,4-mannanase"
 /protein_id="AAC44232.1"
 /db_xref="GI:1491795"
 /translation="MGVCLDMMNRWLVFLVTVIIFLSLIPGTIGFAGVSSVNPVPL
 DFDGFTVMFGAEAGDSLKCIKVSYSDQLRPGNKYALRLDVEFPNNGWDQDGLT
 WFGVYVGQFDFTGYKSFVEFMFIPYDEFKSKQGGPPYKVVINDGWKELGSEFNITAN
 AGKXVINDKDYTVIHKAFAIPEDFTKKRAQLVFOFAGQNSNYKGPVILYDNRVIRPE
 DASLSEYDGGQAEAGVADEDFTGKLVLPQGGKSFVNFELDPMGFKYSGDGN
 KTKYSLEYSQDLKRSNGAALKVNASLGTATFADENMIAIKLTDKDKKDFDKSKYSTLE
 YTVLPNPKISGLKMSAINDPWLIIKDTAINYKDKNAIKNGDKINGDIAVAKCSDN
 LYNNTAKNWLVLRIAGSYKVTGPYIDNLVLGKVAQKVPKVTSSIPKTYFTISST
 KIPAEATKANGVYSIEKENAKYSGKYLILFGNNMGNTLYDLKIPKTYVITISST
 LGNVTSYDVIDTDLGKLAAPVNVKFOEVPVKKVILYTGQETHLSLOKSGGYTI
 AIDYFTIELVLKNKISVDTKLVPNPNPAOKMLKYLASTYKPKILSGQOOSGGD
 KELQMFDTVKRYPAVRGDFDMDYSPSRVEHGKGTGDEEAKNWKSQGIIVAFCHWHN
 APGLIDLPQKGEKGFYEAATFDIKKAMDNPSTEYKILDIRDIDAIIEAQLKRLQKE
 GVPVLPRLHEASGFWGMAQGPYIKLWKLMDLRFRLNYHKLNLDLIIWNGQDAA
 YPDNTVDIIGEDIIYEKAQSYSPADRFKALAKYTNARKMIALTECGTTPDPAVLKQE
 GISWFSFVWAGNVFTVGTSGYNDEWDMHMLRKIYNSDVYVTKDELDPDKNIPLK"
 BASE COUNT 1346 a 630 c 864 g 1060 t
 ORIGIN
 Query Match 6.6%; Score 268.6; DB 1; Length 3900;
 Best Local Similarity 57.6%; Pred. No. 7.8e-69;
 Matches 506; Conservative 0; Mismatches 364; Indels 9; Gaps 1;
 QY 2350 aaacgcttaaccaatccgaatgcagcgttagaagcaaaagcattgatgaactatctggtt 2409
 ||| |||| ||| |||| ||| |||| ||| |||| ||| |||| ||| |||| ||| ||||
 Db 2580 ACAAGGCTTGTAACACCAATCCACATCCTAATGCGCAAAAGCTATGAAATACCTGCCA 2639
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 2410 gatcaatcagggaagaatatgctctctggtcaagag-----gaaataaacgaatt 2460
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 2640 AGCATATGTTGGTGAAGATCTCTCTGCGCAACAGACAGTGTGTGATGGAAGAAAT 2699
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 2461 gattggcttcaagccaatgtaggtataaagccggcattgcagcgctgcacctgatcgac 2520
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 2700 CAATGATATTGATGTTACAAAAAGATATCCTGCTGTCTCAGGGCGTTTGACTTTATGGAC 2759
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 QY 2521 tattgcgaagcagcggaacacggttcttagttccacagaggcagaaaaggcattgca 2580
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 2760 TACTCACCACAGCGGTAGACATGGCACAAAAGGTACTGATGTAGAAGAGGCAATAAAA 2819
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 2591 tgggataagcaagggggatgcttaaccttgcattgacctggaacgcagcagaaaggtctg 2640
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 2820 TGGTGAAGACGGGTGGAAATTTGCTTTCTGCTGGCATTTGAATGACCAACAGGCGTT 2879
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 2641 atcgatcagcagggaagaatggtgagaggtctctatgcgattcaacacacattcgat 2700
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 2880 ATTGACCGCGGCAAGAAATGTTGGAGAGGATTTTACACAGAAGCAACAACTTTGAT 2939
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 2701 atagaatatcgatgaatccagagtcgccgaagattataaattacttcttcgcgacatc 2760
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 2940 ATAAAGAGGCTATGGCAATCCTTAATTCACCTGAGTATAGCTCATTTTGAGAGATATA 2999
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 QY 2761 gatdgattgcaggggcaattgaagaattgcagatgcgaaggttctctgtctctgttcogt 2820
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 3000 GACGCAATTGCTGACGAGCTCAAGAGGCTTCAGAAAAGAGGAGTCCCAAGTCTCTTTTCAGA 3059
 ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||


```

/organism="Ruminococcus flavefaciens"
/strain="FD-1"
/db_xref="taxon:1265"
/cloned="celB"
550..2448
/genes="celB"
550..2448
/genes="celB"
/function="endoglucanase"
/codon_start=1
/transl_table=11
/product="CELB"
/protein_id="AAA95959.1"
/db_xref="GI:736357"
/translation="MRKSTAFILALAMAGCTLSAPLGVLPETKVNAGGDMNLIKVDL
KGERKEISPLIYGVNYTDLKSVKTYAVROGGNRTAYNNENNASNAGSDWKHSSDN
NLSGNPAEVRQRLSKAAYGVDMRTTILMAGIYSADKDGTVKEDVAPSKRWNE
VKTGSAFADPEDLTDGVVYNDENVYIIINLKDSQSPGTIOGYSDENEVPLKNDTH
PRVHPPTVIEELGNKSIELAKAVKLPKAEIFGPALYGYTAFDHLDDDEQVHTESGD
VKSKNNYHWYLDYLDOMKKAEEBEGRLDLDIHYHYSARTGAEDRVOSVRLTYE
EGFSNSWIGOWMCONVPIPTIKKSIDTYVPGTKLAISEVNFKGEDTSGTIAQAEA
LGFADQGVYLATLWGGEPFIISGINLYTNVDKGGCGDFLIPASTEDVSKSTYAA
VNDGESVTVWITNKNMTEAENVIDLENASKDYKSAAYTAVTGNDQVRLDIDVKD
VADKNVVELFASAAWVVSDAAAFDGEKIYBEKKVTEKTEEFKDPSSMINKNGYV
EPIPTDPEQVSKVINGDVTSSAGSNWATARLCSMHQRKGQGFGLDIOELQSPAGQ
GSVRYRKV"
BASE COUNT 725 a 572 c 621 g 574 t
ORIGIN

```

Query Match 4.6%; Score 187; DB 1; Length 2492;

Best Local Similarity 49.7%; Pred. No. 2.7e-44;

Matches 693; Conservative 0; Mismatches 610; Indels 90; Gaps 5;

```

QY 132 tattactattaaagtagatcattcaagatcgtaagcctatagccttataatacgg 191
DB 660 TATGAATATCAAGGTAGACTCAAGGGCGAGGCAAGAGAGATAAGCCCTTATACGG 719
QY 192 tacaatcaggattggcagc-----gatcaaatatggctgccagcagacttggtg 245
DB 720 CGTTACACAGTATACCACAGACCTCAAGAGCGTCAAGACTACTGTGTACGTACAGGGCGG 779
QY 246 caaccgaatgaccggatgacaaatgggaacaaatattgtccaatgaggaagtgcactggca 305
DB 780 TAACCGTATGACTGCTTACAACTGGGAGAACAAATGCTTCAACGCGAGGTTCGGACTGGAA 839
QY 306 gcaatcagcgaataactattatgcagtaattggtggtcgcacaaagcgcgaatgtgaaaa 365
DB 840 ACACAGTTCGGATTAACAACCTCTCGGATTCAAATCCTCTGCTGAGGTA----- 888
QY 366 gccaggagcgtgacgactctttcatgaccaatcgctgaagcttgcaacttattctt 425
DB 889 -----GTTCAGAGACTTTCNANGAAGCTGCAAAAGTACGGCGTTGATTACAAAT 938
QY 426 agttacgttgcgatggcgggttatggctaaaggatggaacggaagtgtgcaggaag 485
DB 939 GACAACCTCCAGATGCGAGGCTACGTTTCAGCTGATAGGAGCGGCACTGTCAAGGAAGA 998
QY 486 cgaaggcccttcctgcgttcgaatcaggtcgtaaacgcgaacaaatgcaccgttcca 545
DB 999 TGAAGTCGCTCTTCAAGCGCTGGAACGAGGTAAAGTTACAAAGGGCGCACCTTTTGC 1058
QY 546 actacagctgactgaatgacaatcggtctatgtgtagatgagttcgctcattttttagt 605
DB 1059 TGATGAGCCTGTACTCACAGACGGCGTGTGTTATATGATGATGATGATGATGATGAT 1118
QY 606 gaacaagtcagcgcactgcttcaacaaggcggggggaaggatgattccctcgacaatga 665
DB 1119 CATAAGCTGGCGATTACAGTACCTACAGGTATCCAGGCTACAGGCTTGATGAAGA 1178
QY 666 accgcctctctggtcgcatacaccacgcattcattgtaaaagtgcgagcgaaga 725
DB 1179 GCCTGTTCTGGAAGGATCTACCCGAGAGTTCATCCGAGAGCTGTAAGTATCGAAGA 1238

```

```

QY 726 gttgtagacgggtcagtcagttattatccaaagctgttaaagcgattacgcgggggcaga 785
DB 1239 GCTTGGAAACAAGTCCATFCGAACCTTGCTAAGGCTGTCAAGAAGCTTGATCTCTAAGGCTGA 1298
QY 786 ggttttggcccggttcttttacggattt-----ggccctataaagatct 830
DB 1299 GATCTTCGACCTGCACCTCTATGGCTACACTGCTTTTCGATCACCCTCGACGATGACGAGCA 1358
QY 831 tcaactgcacctgattgggactctgttaaaaggcaaatatagctggttcgtagactatta 890
DB 1359 GCATACAGATCGGGAGAGCTAAAGTGTCAAGACAACATATCCTGCTGATCTGGACTGCTA 1418
QY 891 cctgattcaaatgcgcttagctcgcaagtcgaagcgaagagatgtgtagattattcga 950
DB 1419 TCTGCACAGATGAAGAAGGCTTCTGAAGAAGAAGGACCCGCTCTCTCGATGATCTG 1478
QY 951 cgtacactgggtatccccaagcgatggcggaagcgcatacgaattacgaataggtagcaa 1010
DB 1479 CATCCACTACTACTCGGAATCTGCAAGAACAGGC----- 1512
QY 1011 tgacgaacgaagaagaccagaatgcagggaacctcgcaaccttggggaccgcgaactataa 1070
DB 1513 -----GCTGAGGACAGAGTTCAGTTCAGTTCGTAATCTCTATGAGGAAGCTTCTC 1562
QY 1071 ggaagatagttggatcgctcaatggaacagcgagtttttgcacatactacactcgatgaa 1130
DB 1563 CGAANAACAGCTGGATCGGTGAGTGATGCAGAAATGTCCTATGAGGACTATATAA 1622
QY 1131 gcagtcggtggataaattatccgggaaccaaagctggcaatgaccgagtagtagctatgg 1190
DB 1623 GAAGTCCATAGATACATACATATATCCGGGCACAAAGCTGGCTATCTCAGAGTATAACTTCAA 1682
QY 1191 cggcgaaatgatatcttcggcggaattggatgacgagatgctggtggtatctctgggcaa 1250
DB 1683 GGGCGGCGAGGATACTTCCGGTACTATCGCTTCAGGCGAGAGGACACTGGGATGCTTCGGCTGA 1742
QY 1251 aaatgatgtttatgacaaactactggaagcgaagcgaagtggtgtcaacaactacgttag 1310
DB 1743 TCAGGGCGTATATCTCGCACTCTCTGG-----GGCGGTGAGCCATTCATCAT 1790
QY 1311 tgcgcttacaggttttactcgcaattatgacggaacaaactcttcttcggtgataccag 1370
DB 1791 CTCAGGTATCAACCTCTACACTAACTACGAGCGCAAGGGCGGATGCTTCGGTGATACTCT 1850
QY 1371 tgttagtgccgaacacatcgatattgtcaatagctcgttcgctcattctgtaacgaatgc 1430
DB 1851 CATCCCTGCTTCTACTGAAAGTGTATCAAAAGTCCAGCACACATACGCTGCTGTTAAACGCG 1910
QY 1431 atccgacaaagaaactcgatctcgttgcataataaagcatgacagcgattcgcgcgc 1490
DB 1911 CGATGAGTCAAGGTAACTGTCATGTATGATCACCACAAAGTAATGACAGAAGCTGAGAAGC 1970
QY 1491 ccaatttgatctt 1503
DB 1971 TGTTATCGACCTT 1983

```

RESULT 15

RFL298117 2692 bp DNA linear BCT 10-OCT-2001
LOCUS Ruminococcus flavefaciens 17 endB gene for cellulase.

DEFINITION RFL298117

ACCESSION AJ298117.1 GI:16040919

VERSION AJ298117.1

KEYWORDS cellulase; endB gene.

SOURCE Ruminococcus flavefaciens.

ORGANISM Ruminococcus flavefaciens

Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

Ruminococcus.

REFERENCE 1 (bases 1 to 2692)

AUTHORS Rincón, M.T.

JOURNAL Thesis (2001); Department of Agriculture, University of Aberdeen,

Aberdeen, United Kingdom

```

REFERENCE 2 (bases 1 to 2692)
AUTHORS Rincon,M.T., McCrae,S.I., Kirby,J., Scott,K.P. and Flint,H.J.
TITLE EndB, a Multidomain Family 44 Cellulase from Ruminococcus
flavifaciens 17, Binds to Cellulose via a Novel Cellulose-Binding
Module and to Another R. flavifaciens Protein via a Dockerin Domain
Appl. Environ. Microbiol. 67 (10), 4426-4431 (2001)
MEDLINE 21454991
REFERENCE 3 (bases 1 to 2692)
AUTHORS Flint,H.J.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2000) Flint H.J., Gut Microbiology Group, Rowett
Research Institute, Greenburn Road, Bucksburn, Aberdeen, AB21 9SB,
UNITED KINGDOM
FEATURES
source
1. .2692
Location/Qualifiers
/organism="Ruminococcus flavifaciens"
/strain="17"
/db_xref="taxon:1265"
/country="United Kingdom"
/length="271"
/gene="endB"
/feature="endB"
/function="beta-(1,4)-glucanase"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="cellulase"
/protein_id="CAC83072.1"
/db_xref="GI:16040920"
/translation="MKRTAFIAACVSVSGCTWTPVNGLPRAITVNAAGGYDMNVTVDL
KGKKAIPLIYGVNYTDLDRVKTAVRQGNRTAYNWETNASNGSDWKHSSDN
NLSDSDPDCVQVLGSKQAKYVNYKLITLQLAGVSDAKNGPVSEAEKAPSDRWNK
VLTKNAPDFTPLDGLVYVDEYVNYIINKLSDSAGIQGYSLDNEPVLWNTDTH
SRMHPDPTVEELGSKSVEMAKVKILDPKAEVFGPALYGTAFQHLDDDDAHTWEE
IKNNYHWLDCYLDHMHKASEENARLLDLDIHYSESARKGIEDRLQSVRLIYE
PGSENSWQWCWENPILPTOKSIDITYPGTKLIGISEYNFEGGDDASGTIAQAE
LGYADQGVTFASLWGEPEFILSGIQLYNYDGGCGFGDTLIPASTGDVSKSSTYAA
VNAKDSKVTVYNTKDLKENENAVIDLKRNADSKYSAAYAVFVSESEIRLIDIKD
VKNKVTLELPASAAVYVSDQADFGLKTYEETKTKTVEFTDIESMNDKGFV
VVPIDAEHLKSIINGAVTSSAGSSWATAGCAVMNVTAKDGSFGFTVYSNLPGLS
KSAATVFDGIFNKEGEGADKVKEDLEATVAGKVELOKWDASEKSESEDEKIEV
KYSIQVYIYAGAEAPKVTITTSNAVTTTTSVATTTSSATPDKDVYVGDANDCG
NVDLADAILTQSLANPKNFGLNGSAEKHLTEKGRNADCCDPGSLITNDLALSQRL
LLHLIDLPE"
176..2602
/gene="endB"
272..2599
/gene="endB"
/mat_peptide
721 a 683 c 724 g 564 t
BASE COUNT 721 a 683 c 724 g 564 t
ORIGIN

Query Match 4.43; Score 178.6; DB 1; Length 2692;
Best Local Similarity 49.54; Pred. No. 9.4e-42;
Matches 709; Conservative 0; Mismatches 634; Indels 90; Gaps 6;

Qy 55 gtcgtcagtcgacgatagtagtggggggtttacctgctccagctggtgttcac 114
Db 209 GTAGTATCGGCTGCACAAAGACAGCGCCGCTCAACGGTGTGCGAGGGGTACCGTTAAT 268
Qy 115 ggtcaaacgcaagaactattactattaaagtagatcacattcaaggatcgtaagcctatt 174
Db 269 GCGGCGGGAGGCTACGACATGAACCTGACCGTAGACCTCAAGGTGAGAAAAAGGCTATC 328
Qy 175 agcccttatatacgttacaaatcaggatttggcag-----gcgatgaaatatggct 228
Db 329 TCTCCTCTTACTACGGCGTGAACCATATACACCCAGCTTACGGGACGTTAAGACTACG 388
Qy 229 gccagacgacttgggcaacccgaatgacgggatacacactgggaaacaaatattgtccaat 288
Db 389 GCTGTGCGTCAGGCGGTAAACCGATGACTGCGCTACAAATTGGGAGACAAATGCTTCAAT 448

```

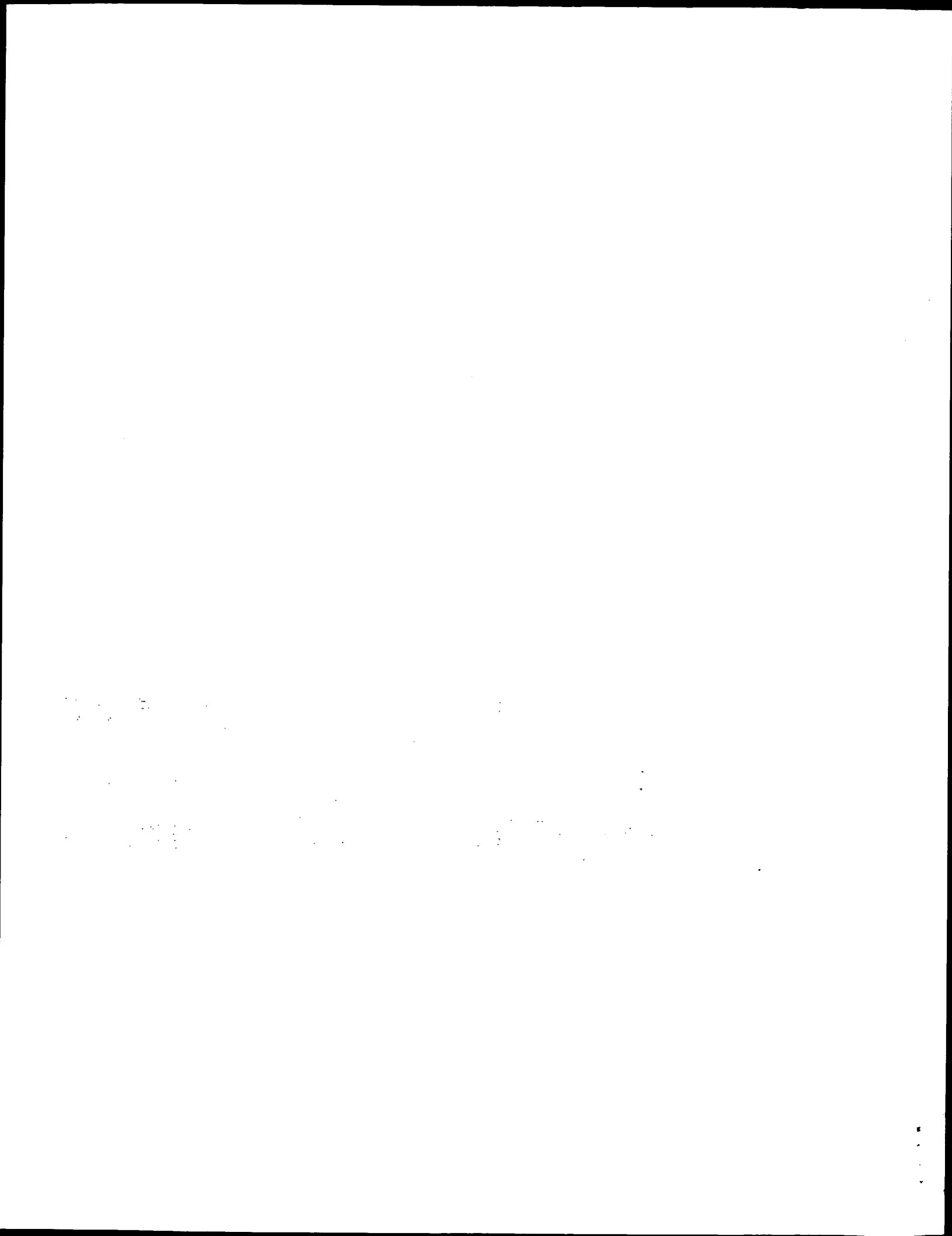
```

QY 289 gcaggagtgactgggcagcaatctagcgataaactattatgcagtaattggtgcctcgaca 348
Db 449 GCGGGTTCGACTGGGAAGCACAGCTCGGCACACAACCTCTCGATTCGATGATCCTGGG 508
QY 349 caagccgaatgtaaaagccagagcggtgagcactctgttcaatgaccaatcgctgaag 408
Db 509 GACTGCGTTTCAGTGCTCTCAAAAGCAGCGGGTAAAGTACAAATGTAACCTACAAAGTCA-- 566
QY 409 cttggcaattattctttagtttagcttggcgggttatgtggttaagatggaac 468
Db 567 -----CTACCCCTCCAGCTTGCAGGCTATGTTTACGCTAGTATAAGAAC 607
QY 469 ggaagtgtgcagaaagcaaaagggcccttcgctggttggaatcagctcgtaaacgcc 528
Db 608 GGCCTCTGTTTCCGAGCGGGAAGAGGCTCCTCTGACCCGCTGGACAACAGGTGCTGTACA 667
QY 529 aaaaatgcacggttccaaactacagcctgatctgaaatgacaaatcggggtctatgtgatgag 588
Db 668 AAGAAATGCTCCCTTTGCGGATATCTCCGACCTCACAGACGGAGTGTCTTACATGACGAG 727
QY 589 ttgcctcatttttagtgaaacagtcagcactgcttcaacaaagcgcggtggaagga 648
Db 728 TAGTGAACTACATATAAATAAGCTGGCGGATTCACAGTCGGCTGAGGGTATCCAGGGC 787
QY 649 tatgcctcgacaatgaacccgctctctgctgctacacgacccacgcaattcatggtgaa 708
Db 788 TAGAGCTTGACAATGAGCCCGTCTCTGGAACGATACACACAGCAGGATGACCCCTGAT 847
QY 709 aaagtgcggcgaaagagttgtagacgggtcagtcagtttatccaaagctgtgaaagcg 768
Db 848 CCGGTTTACTATGAGGAGCTGGCGAGCAATCCGTTGAAATGGCAAGAGCTGTAAAGAG 907
QY 769 attgacgcggggcgagagatttttggccggtctcttaccgatttggcgctataaagat 828
Db 908 CTGGATCCTAAGGCTGAGGTGTTGCGCCCGCTCTCTACGGCTACACCGCCTTCGATCAC 967
QY 829 ctcca-----aactgcacctatggggactctgtaaa-----aggcaattatagc 873
Db 968 CTTGACGAGGATGACGCTCATACCGAGTGGGAGGAGATAAAGAGGCCAATAATTACCAT 1027
QY 874 tggttogtagactattaccctgagcaaatcgcccttagctcgaagtgcgaagcgaaagaga 933
Db 1028 TGGTATCTGGAGTGTCTACCTCTCATATGTCACAAAGGCTCCGAGGAGAACGGCGCAAGA 1087
QY 934 ttgctgatgtattgacgtacactggttatcccgaaagcgatggcgagcgatacgaatt 993
Db 1088 CTTCTTGAGCTTCTGGATATACATATATTATTCGGAGTACAGCAAGAAAGGGCATCG- 1142
QY 994 acgaatgaggtaggcaatgacgaaacgaagaagcagaatgcaggcaccctgcaccttg 1053
Db 1143 -----AGGACAGGCTCCAGTCTGCTGCGCACACTC 1171
QY 1054 tgggaccgcacataaaggaagatagttggtatcgctcaatggaacagcgagtttttgccc 1113
Db 1172 TATGAGCCCGGTTCTCAAAAACACAGTGGATAGTGTAGTGTGATGGAGAACCTGGCG 1231
QY 1114 atactacctcgattgaagcagtcggtggaataatattccgggaaacaaagctggcaatg 1173
Db 1232 ATACTTCTTACCATACAGAGTCCATCGATACTTACTATCTTGGCACAAAGCTGGGCATA 1291
QY 1174 accgagtagtctatggtggcgcaaaatgatatcttcggcgggatgctgcatgaccgatg 1233
Db 1292 TCCGAGTACAACTTCCGCGCGCGACGATGCTTCCGGGACATATCGCTCAGGCTGAGGCT 1351
QY 1234 ctgggttatctgggcaaaaatgatgtttatatggaacactactggaagcctaaagatggt 1293
Db 1352 CTGGGCTGTATCGGATCAGGGCGGTGTACTTTGCTTCCTCTGG-----GGC 1399
QY 1294 gtcaacaactagctagttagtcgctcttacaagctttatcgcaattatgacgggaaacactct 1353
Db 1400 GCGGAGCCCTTTATCTCTCGGGTATCCAGCTCTATACCACTATGACCGCAAGGCGGC 1459

```

Qy 1354 actttcgtgataccagtggttagtgcgcgaacatcgcgatatgtcaatagctcgggccat 1413
Db 1460 TGCTTCGGCGATACCCCTTATCCCTGCATCTACCGGGGATGTTCCCAAGTCCAGCACCTAT 1519
Qy 1414 gcttctgtaacgaatgcataccgcacaaagaactgcatactcgtttgcatgaataa 1466
Db 1520 GCGGCTGTAAACGCAAGGAGGATTCAAAGGTCACCTGTATGTCACAAACAA 1572

Search completed: July 12, 2002, 18:45:14
Job time: 32701 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 10:25:23 ; Search time 1513.98 Seconds
(without alignments)
4603.070 Million cell updates/sec

Title: US-09-784-554B-1

Perfect score: 4059

Sequence: 1 atgaggcgcaaaatagtag.....ccgagcgacagctaattaa 4059

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4059	100.0	4059	22	AAH75059 Nucleotide sequenc
2	3374	83.1	4056	22	AAH75060 Nucleotide sequenc
3	1653.4	40.7	2141	22	AAH75061 Nucleotide sequenc
4	602.2	14.8	2977	12	AAQ13001 Endol gene encodin
5	601.4	14.8	1854	12	AAQ14842 B.lautus endol cor
6	431.6	10.6	2445	21	AAZ45347 DNA encoding a Bac
7	405.2	10.0	6415	20	AAZ55662 DNA sequence encod
8	405.2	10.0	6416	24	AAZ26526 Active cellulase p
9	404.2	10.0	1488	21	AAZ45348 DNA encoding a Bac

10	333.2	8.2	1464	21	AAZ45341 DNA encoding a Hum
11	261.8	6.4	3041	21	AAZ45350 DNA encoding a map
12	144.6	3.6	2600	12	AAQ15178 Portion of pAEC-1
13	93.2	2.3	11706	20	AAZ55661 DNA sequence encod
14	93.2	2.3	11707	24	AAZ26525 Active cellulase p
15	87.8	2.2	2029	20	AAZ55660 DNA sequence encod
16	87.8	2.2	2029	24	AAZ26568 Active cellulase h
17	87	2.1	2175	14	AAQ49820 NK-1 cellulase gen
18	68.2	1.7	1775	12	AAQ13003 Endo3 gene encodin
19	50.8	1.3	1461	10	AAZ91621 Beta-mannase gene
20	50.6	1.2	4839	11	AAQ06827 Alpha amylase pull
21	45.8	1.1	5059	20	AAZ84332 Stealth virus nucl
22	44.6	1.1	2892	18	AAZ92376 Cyclic-isonaltolig
23	43.4	1.1	1086	21	AAZ45349 DNA encoding a Bac
24	40.4	1.0	1011	21	AAZ29313 B. subtilis strain
25	40.4	1.0	1029	21	AAZ29847 Bacillus subtilis
26	40	1.0	1687	20	AAV84559 Human secreted pro
27	40	1.0	1687	22	ABA83342 Human secreted pro
28	40	1.0	1693	21	AAZ59175 Human secreted pro
29	38.2	0.9	1438	21	AAZ45336 DNA encoding a man
30	38.2	0.9	1482	20	AAZ90978 DNA encoding pecta
31	38.2	0.9	1482	20	AAZ31562 Pectate lyase CBD
32	38.2	0.9	5562	19	AAZ86625 C. thermocellum Ci
33	38	0.9	2916	17	AAZ15957 Cycloisomaltoligo
34	38	0.9	2916	20	AAZ10902 Bacillus Citase co
35	37.4	0.9	2984	23	AAZ86393 DNA encoding novel
36	37.4	0.9	14041	22	AAH48024 Internal control B
37	37.2	0.9	1030	22	AAZ11044 Clostridium cellul
c	38	0.9	22008	22	ABA15839 Human nervous syst
39	37	0.9	22008	22	AAZ36610 Human cardiovascular
c	40	0.9	783	21	AAZ53635 Neisseria meningit
c	41	0.9	783	21	AAZ53636 Neisseria meningit
c	42	0.9	798	21	AAZ53634 Neisseria gonorrh
c	43	0.9	9905	24	ABL32062 Human immune syste
c	44	0.9	44608	21	AAA81495 N. meningitidis pa
45	36.4	0.9	349980	21	AAZ21607 Neisseria meningit

ALIGNMENTS

RESULT	1
AAH75059	
ID	AAH75059 standard; DNA; 4059 BP.
XX	
AC	AAH75059;
XX	
DT	29-OCT-2001 (first entry)
XX	
DE	Nucleotide sequence of xyloglucanase enzyme.
XX	
KW	xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW	cellulosic fiber; textile scouring; ss.
XX	
OS	Paenibacillus polymyxa.
XX	
FH	Key Location/Qualifiers
FT	CDS 1..4059
FT	/*tag= a
FT	/product= "xyloglucanase"
XX	
PN	WO200162903-A1.
XX	
PD	30-AUG-2001.
XX	
PF	21-FEB-2001; 2001WO-DK00116.
XX	
PR	24-FEB-2000; 2000DK-0000291.
XX	
PA	(NOVO) NOVOZYMES AS.
XX	
PI	Schnorr K, Jorgensen PL, Schuelein M;
XX	

DR WPI: 2001-522819/57.
DR P-PSDB: AAG63962.

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries

XX Claim 14; Page 80-81; 97pp; English.

XX The present sequence encodes a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees celsius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for ratting of fibers e.g. hemp, jute, flax, and
XX linen.

SQ Sequence 4059 BP; 1187 A; 893 C; 1046 G; 933 T; 0 other;

Query Match 100.0%; Score 4059; DB 22; Length 4059;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4059; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagggcgaaatagtagtaatttttttgcacacgcttcacaaatgctgcctgtcgc 60
Db 1 atgagggcgaaatagtagtaatttttttgcacacgcttcacaaatgctgcctgtcgc 60
Qy 61 atggcttcacgagtagtagtagggggctttacccttcacgctggttcacggtcaa 120
Db 61 atggcttcacgagtagtagtagggggctttacccttcacgctggttcacggtcaa 120
Qy 121 acggcaaaagactattactattaaagtagatgacattcaagtagctgaagcctattagccct 180
Db 121 acggcaaaagactattactattaaagtagatgacattcaagtagctgaagcctattagccct 180
Qy 181 tatatatcggtacaaatcaggatttggcaggcgatgaaataatggctgcagacgactt 240
Db 181 tatatatcggtacaaatcaggatttggcaggcgatgaaataatggctgcagacgactt 240
Qy 241 ggtggcaaccgaatgacggatgacacactgggaaacaaatattgtccaatgacgaagtgc 300
Db 241 ggtggcaaccgaatgacggatgacacactgggaaacaaatattgtccaatgacgaagtgc 300
Qy 301 tggcagcaatctagcagataactatttatgagtagtaattggtggcctgacacaaagccgaatgt 360
Db 301 tggcagcaatctagcagataactatttatgagtagtaattggtggcctgacacaaagccgaatgt 360
Qy 361 gaaagccagagcggtagcagacttcgttccatgaccaaactgcgtgaagcttggcacttat 420
Db 361 gaaagccagagcggtagcagacttcgttccatgaccaaactgcgtgaagcttggcacttat 420
Qy 421 tcttagttacgttgcgagtagccggttatgtggttaagtaggaaacgaagtagtgcag 480
Db 421 tcttagttacgttgcgagtagccggttatgtggttaagtaggaaacgaagtagtgcag 480
Qy 481 gaaagcgaaagggcccttcgcctcgttggaaatcagctgtaaacgcccacaaatgacccg 540
Db 481 gaaagcgaaagggcccttcgcctcgttggaaatcagctgtaaacgcccacaaatgacccg 540
Qy 541 ttcaactacagcgtgatctgaatgacaaactcggtcttatgtgatgagttcgctccatttt 600
Db 541 ttcaactacagcgtgatctgaatgacaaactcggtcttatgtgatgagttcgctccatttt 600
Qy 601 ttagtgaacagtagtcgcactgcttcacaaagcggggtgaaagatgacccctgcag 660
Db 601 ttagtgaacagtagtcgcactgcttcacaaagcggggggtgaaagatgacccctgcag 660

Qy 661 aatgaaccgcgtctctggtcgcatacgacccacgcatctcatgttgtaaaaaagtcggagcg 720
Db 661 aatgaaccgcgtctctggtcgcatacgacccacgcatctcatgttgtaaaaaagtcggagcg 720
Qy 721 aaagagttgtagacacgggtcagtcagtttatccaaagctgtgaaagcagattgacgcgggg 780
Db 721 aaagagttgtagacacgggtcagtcagtttatccaaagctgtgaaagcagattgacgcgggg 780
Qy 781 gaagagtttttggcccggtctcttaccgatttggcgctataaaagattctcaaacctga 840
Db 781 gaagagtttttggcccggtctcttaccgatttggcgctataaaagattctcaaacctga 840
Qy 841 cctgattgggactctgtaaaaggcaattatagctggttgcgtagactattacctggatcaa 900
Db 841 cctgattgggactctgtaaaaggcaattatagctggttgcgtagactattacctggatcaa 900
Qy 901 atgccccttagctgcgaagtcgaagcgaagagattgctggtatgatttcgacgtacactgg 960
Db 901 atgccccttagctgcgaagtcgaagcgaagagattgctggtatgatttcgacgtacactgg 960
Qy 961 tatccgaaagcagtagggcgagggcagacacgaattacgaatgaggtaggtgacgaacacg 1020
Db 961 tatccgaaagcagtagggcgagggcagacacgaattacgaatgaggtaggtgacgaacacg 1020
Qy 1021 aagaagccagaaatgcaggcacctcgcaccttctgtggaccccgacacctataaagaagtagt 1080
Db 1021 aagaagccagaaatgcaggcacctcgcaccttctgtggaccccgacacctataaagaagtagt 1080
Qy 1081 tggatcgctcaatggaaacggcgagtttttgcacatactactacctcgattgaagcagtcgg 1140
Db 1081 tggatcgctcaatggaaacggcgagtttttgcacatactactacctcgattgaagcagtcgg 1140
Qy 1141 gataaataattatccgggaaacaaagctgcgaatgacacgaatgacgtatgctgacggcgaat 1200
Db 1141 gataaataattatccgggaaacaaagctgcgaatgacacgaatgacgtatgctgacggcgaat 1200
Qy 1201 gatattccggcggtgattgcgtagcagatgctggtggttatcttgggcaaaaatgatgtt 1260
Db 1201 gatattccggcggtgattgcgtagcagatgctggtggttatcttgggcaaaaatgatgtt 1260
Qy 1261 tataatgcaactactggaagctaaaggtggtgtcaacaaactacgttagtgcgccttac 1320
Db 1261 tataatgcaactactggaagctaaaggtggtgtcaacaaactacgttagtgcgccttac 1320
Qy 1321 aagctttatcgcaattatgacgggaaaaaactctacttctcgggtgacacaggttagtgcg 1380
Db 1321 aagctttatcgcaattatgacgggaaaaaactctacttctcgggtgacacaggttagtgcg 1380
Qy 1381 caaacatcggtattgtcaatagctcgttccatgcttctgttaacgaatgcatccgacaaa 1440
Db 1381 caaacatcggtattgtcaatagctcgttccatgcttctgttaacgaatgcatccgacaaa 1440
Qy 1441 gaactgcatctcgttgcataaaagcattggacagcgaatcgcagcccaatttgcg 1500
Db 1441 gaactgcatctcgttgcataaaagcattggacagcgaatcgcagcccaatttgcg 1500
Qy 1501 ctttcggcgcaagaacttaccatttcccgtaaaagtagtgggttcgataaaacagctcg 1560
Db 1501 ctttcggcgcaagaacttaccatttcccgtaaaagtagtgggttcgataaaacagctcg 1560
Qy 1561 caaattaaagaagcagcgaatcgcgaatcgcgaatcgcgaatcgcgaatcgcgaatcgcga 1620
Db 1561 caaattaaagaagcagcgaatcgcgaatcgcgaatcgcgaatcgcgaatcgcgaatcgcga 1620
Qy 1621 ccgctttgacggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680
Db 1621 ccgctttgacggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680
Qy 1681 ggtcctgaaagctttaagctgaaagctgaggtggtgagtgaggaaagtcacatttaccctgg 1740
Db 1681 ggtcctgaaagctttaagctgaaagctgaggtggtgagtgaggaaagtcacatttaccctgg 1740

QY 1741 gatcttccagcgaggttaggatagacagctacagcgtagcgagcgcaacagatgaaacgcccct 1800
Db 1741 gatcttccagcgaggttaggatagacagctacagcgtagcgagcgcaacagatgaaacgcccct 1800
QY 1801 ttaactgcttagatcccaactgacgaaacgtcttatacagataactaactgacgaaac 1860
Db 1801 ttaactgcttagatcccaactgacgaaacgtcttatacagataactaactgacgaaac 1860
QY 1861 ggtacttactattacaagtaaacgccaacaaacaaataaaggatcgagcgaaatccaat 1920
Db 1861 ggtacttactattacaagtaaacgccaacaaacaaataaaggatcgagcgaaatccaat 1920
QY 1921 atttgaagcggttcgaagatgctgttaaacggtcccgctcgctataaagccgaagaa 1980
Db 1921 atttgaagcggttcgaagatgctgttaaacggtcccgctcgctataaagccgaagaa 1980
QY 1981 ggcacgctgaaggaacacatttggaatccagcgagcggtactccggtgctggttat 2040
Db 1981 ggcacgctgaaggaacacatttggaatccagcgagcggtactccggtgctggttat 2040
QY 2041 gtaacgaattccacaatccaggggattctctgacgatgacgattcaggtccacggca 2100
Db 2041 gtaacgaattccacaatccaggggattctctgacgatgacgattcaggtccacggca 2100
QY 2101 ggtctgtacaatcttacaatcggtactcggttctctctcatgatgacaaacgacccaatttc 2160
Db 2101 ggtctgtacaatcttacaatcggtactcggttctctctcatgatgacaaacgacccaatttc 2160
QY 2161 tcaataaacggaagcggttggcgaactgctgttaaaaaacggtgattttaagaa 2220
Db 2161 tcaataaacggaagcggttggcgaactgctgttaaaaaacggtgattttaagaa 2220
QY 2221 acttccgaggaaggtgctgtggaatgagcgcggaatacagatcaggtttgaaacaggc 2280
Db 2221 acttccgaggaaggtgctgtggaatgagcgcggaatacagatcaggtttgaaacaggc 2280
QY 2281 tggggctggttacgatactgactacgtcagactgagcggtcgctcgctgacgcgccacatcat 2340
Db 2281 tggggctggttacgatactgactacgtcagactgagcggtcgctcgctgacgcgccacatcat 2340
QY 2341 gcggtacccaaacgcttaccatccgaatgcgacggttagaagcaaaagcatttgtaac 2400
Db 2341 gcggtacccaaacgcttaccatccgaatgcgacggttagaagcaaaagcatttgtaac 2400
QY 2401 tatctggttatcaatacgggaagaatactctctgttcaagaggaataaaacgaatt 2460
Db 2401 tatctggttatcaatacgggaagaatactctctgttcaagaggaataaaacgaatt 2460
QY 2461 gattggcttcaagccaatgtagttaaaacgcggtgattgacgctgacgtgacgtgac 2520
Db 2461 gattggcttcaagccaatgtagttaaaacgcggtgattgacgctgacgtgacgtgac 2520
QY 2521 tattcgccaagcagagcggaacacggtcttagttccacagagcgagaaaggcgattgca 2580
Db 2521 tattcgccaagcagagcggaacacggtcttagttccacagagcgagaaaggcgattgca 2580
QY 2581 tgggataagcaagggggatcggttacacttgcagtgcaactggaacgacgaaaggtctg 2640
Db 2581 tgggataagcaagggggatcggttacacttgcagtgcaactggaacgacgaaaggtctg 2640
QY 2641 atcgatacgcagggaaagaatggtgagaggtctctatcggtatcaaccacattcgat 2700
Db 2641 atcgatacgcagggaaagaatggtgagaggtctctatcggtatcaaccacattcgat 2700
QY 2701 atagaatatcgatgaatccagagtcggaagattataaattactatttcgcgacatc 2760
Db 2701 atagaatatcgatgaatccagagtcggaagattataaattactatttcgcgacatc 2760
QY 2761 gatgtgattcagggcaattgaagaattgcaaggtgcaaggttccgttccgttccgt 2820
Db 2761 gatgtgattcagggcaattgaagaattgcaaggtgcaaggttccgttccgttccgt 2820
QY 2821 cctttgacgaagcggaagaaaatggttctggtggggcgccaaggctcgtgagcctggt 2880

Db 2821 cctttgacgaagcggaagaaaatggttctggtggggcgcaaaaggctcgtgagcctgtt 2880
QY 2881 aaaaagctgtatatatttaatacagacacggtttgacgaatgtgcacaaattgacaaatctg 2940
Db 2881 aaaaagctgtatatatttaatacagacacggtttgacgaatgtgcacaaattgacaaatctg 2940
QY 2941 atttgggtatggaattctgttgcctccggtatggtatccgggagacgagatattggtgattt 3000
Db 2941 atttgggtatggaattctgttgcctccggtatggtatccgggagacgagatattggtgattt 3000
QY 3001 ttgagctttgactcttattccgcaagcaggtgatttacagcccgcaaaatttcaaaatacga 3060
Db 3001 ttgagctttgactcttattccgcaagcaggtgatttacagcccgcaaaatttcaaaatacga 3060
QY 3061 gaccttggctgattgggcaagacaaaagctagttgcatgagcgaaaatggaccgac 3120
Db 3061 gaccttggctgattgggcaagacaaaagctagttgcatgagcgaaaatggaccgac 3120
QY 3121 ccggaacctgatttgatgaagcggtatcaagctcatgagctggttgcgtacatggtat 3180
Db 3121 ccggaacctgatttgatgaagcggtatcaagctcatgagctggttgcgtacatggtat 3180
QY 3181 ggagattttgagagcgcaaaacaaacagccttgagcatctgaaaaagtgataat 3240
Db 3181 ggagattttgagagcgcaaaacaaacagccttgagcatctgaaaaagtgataat 3240
QY 3241 catcgaaogtcttaactgagctgagctcccaacgaacttaaaacgtatggcattact 3300
Db 3241 catcgaaogtcttaactgagctgagctcccaacgaacttaaaacgtatggcattact 3300
QY 3301 gacgacgctcogtaccggcgagcttcaagctgaacgctgaggtggaacgcaaaagta 3360
Db 3301 gacgacgctcogtaccggcgagcttcaagctgaacgctgaggtggaacgcaaaagta 3360
QY 3361 tgcgttaagctgacagcatcgcggaatgcgaaagctatgaagtgaagcttcgcagact 3420
Db 3361 tgcgttaagctgacagcatcgcggaatgcgaaagctatgaagtgaagcttcgcagact 3420
QY 3421 gaaaacgcgcttcgcaactgtagcagtgatgtatatggcagtagctacacagacaca 3480
Db 3421 gaaaacgcgcttcgcaactgtagcagtgatgtatatggcagtagctacacagacaca 3480
QY 3481 gctgtaaagcgagatagcagctactacaaagcggtatgagcaagacagatgcagagacag 3540
Db 3481 gctgtaaagcgagatagcagctactacaaagcggtatgagcaagacagatgcagagacag 3540
QY 3541 acgctgtcgaaacacgcttagcgcaatgcgaaagcggtatctacagcagcgacagacagga 3600
Db 3541 acgctgtcgaaacacgcttagcgcaatgcgaaagcggtatctacagcagcgacagacagga 3600
QY 3601 ctgctgtctccaatatcgcaacagcagacttaacgatacagatcaacagattgaaattcgctac 3660
Db 3601 ctgctgtctccaatatcgcaacagcagacttaacgatacagatcaacagattgaaattcgctac 3660
QY 3661 ttccaaattgttaaaacaaaggcaacacctccataccgatacagcagattgaaattcgctac 3720
Db 3661 ttccaaattgttaaaacaaaggcaacacctccataccgatacagcagattgaaattcgctac 3720
QY 3721 tactacaaatcgagcggtgacgctgagcagactttcaactgcgactatgcgacgctgagc 3780
Db 3721 tactacaaatcgagcggtgacgctgagcagactttcaactgcgactatgcgacgctgagc 3780
QY 3781 tgcctcaagctgaacggttaaaactggttaaaatggagaagctgcacccggtgcgattat 3840
Db 3781 tgcctcaagctgaacggttaaaactggttaaaatggagaagctgcacccggtgcgattat 3840
QY 3841 tatttgggaagtcagtttaacttcgagtcagcggtgttagcacctggagggaagcagcggc 3900
Db 3841 tatttgggaagtcagtttaacttcgagtcagcggtgttagcacctggagggaagcagcggc 3900
QY 3901 gatatccaaacccggtattcataagacagactggtcgaactataacgaagtgacgattat 3960

Db 3901 gatataccaaacccgtattcataagacagactggtcgaaactataacgaaagtgcgatttat 3960

Qy 3961 tcgtataaggcagcgaacacctcatttcccgatcatctcaatgaattaccttgatcataac 4020

Db 3961 tcgtataaggcagcgaacacctcatttcccgatcatctcaatgaattaccttgatcataac 4020

Qy 4021 ggcgtacttgttgggaacccgagccgacagctaataa 4059

Db 4021 ggcgtacttgttgggaacccgagccgacagctaataa 4059

RESULT 2

AAH75060

ID AAH75060 standard; DNA; 4056 BP.

XX

AC AAH75060;

XX

DT 29-OCT-2001 (first entry)

XX

DE Nucleotide sequence of xyloglucanase enzyme.

XX

DE xyloglucanase; family 44; glycosyl hydrolase; detergent;

KW cellulose fiber; textile scouring; ss.

KW

XX Paenibacillus polymyxa.

OS

XX

FH Key

FT CDS

FT 1..4056

FT /*tag= a

FT /product= "xyloglucanase"

FT /transl_except= "(pos: 4039..4044, aa: Glu)"

XX

PN WO200162903-A1.

XX

XX 30-AUG-2001.

XX

XX 21-FEB-2001; 2001WO-DK001116.

XX

PR 24-FEB-2000; 2000DK-0000291.

XX

XX (NOVO) NOVOZYMES AS.

PA

XX

XX Schnorr K, Jorgensen PL, Schuelein M;

PI

XX WPI; 2001-522819/57.

DR

XX P-PSDB; AAG63963.

XX

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family.

PT useful for detergent compositions, and textile or cellulose fiber

PT processing industries

XX

PS Claim 14; Page 85-86; 97pp; English.

XX

XX The present sequence encodes a xyloglucanase of the invention. The

CC specification describes a xyloglucanase enzyme belonging to family 44

CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity

CC of at least 30% at pH 5-8. The enzyme exhibits high performance in

CC detergent compositions and prevents binding of certain soils to the

CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at

CC room temperature and has a half life of more than 50 days when incubated

CC in a full formulated liquid detergent at 30 plus degrees celsius. The

CC enzyme is used in detergent compositions, textile industry for improving

CC the properties of cellulosic fibers, yarn, and woven or non-woven

CC fabrics, preferably in textile scouring process, and in cellulose fiber

CC processing industry for rattling of fibers e.g. hemp, jute, flax, and

CC linen.

XX

SQ Sequence 4056 BP; 1161 A; 888 C; 1070 G; 935 T; 2 other;

Query Match 83.1%; Score 3374; DB 22; Length 4056;

Best Local Similarity 90.2%; Pred. No. 0;

Matches 3664; Conservative 2; Mismatches 387; Indels 9; Gaps 5;

Qy 1 atgaggcgcaaaaatagtagtaattctttgttcaaacgttccaaatggctgctgcgtc 60

Db 1 atgaagcgcaaaaatagtagtagtatttgggtccaaacgttccaaatggctgctgcgtc 60

Qy 61 atggcctgcacgatgatagtaggggggttttacctgctccagctgtggttcaacgtcaa 120

Db 61 atggcatgcagattatagtaggggggtgctctaccgactccaactgtagtctacggctcaa 120

Qy 121 acggcaagactattactattaaagttagatatactcaagatcgtaagcctattagccct 180

Db 121 acggcaagactgttaccattaaagtcgatactcaagatcgtaagcctattagccct 180

Qy 181 tatatacgggtacaaaatcaggatttggcaggcgatgaaataatggtctccagagacct 240

Db 181 tatattacggtacgaatcaggagttggcaggcgatgagaatctgactccagagacct 240

Qy 241 ggtgcaacccgaatgacgggatacaactgggaaaaacaatatgtccaatgcagaaagtgc 300

Db 241 ggtgcaatcgaaatgacgggatacaactgggaaaaacaatatgtccaatgcagaaagtgc 300

Qy 301 tggcagcaatctagcgataactatttatcagtaataatgggtggcctgacacaaacgaatg 360

Db 301 tggatgcagtcacagcgatagctatttatcgacaacacgcccggatttgacaaaacccaagt 360

Qy 361 gaaaagccaggcggtgacgactgttcttcacgaccaatcgctgaagcttgaccttat 420

Db 361 gaaaagccagggtgcggtggcaacctgtttcagatcaatcgctgaagcaggacacatat 420

Qy 421 tctttagttacgtgcccgtatgcccgggttatgtggcctaagatggaacggaaagtgcag 480

Db 421 tctttagtcacactgcccgtatgcccgggttatgtggcctaagatggaacggaaagtgcag 480

Qy 481 gaaagcgaagggcccttccgctgtggaatcaggtcgtaaacgcaaaaatgcacgc 540

Db 481 gaaagcgaagggccttccgctgtggaatcaggtcgtaaacgctaaaatgcgcgc 540

Qy 541 ttccaaactacagcctgactgactgaatgacaaatcggtgtctatgtgtgagtgagttccatttt 600

Db 541 ttccaattgcagcctgactgactgaagacaaatcaggtttatgcggatgaatctgcaacttt 600

Qy 601 ttatgaaacaagtacggcactcttcaacaaaggcgggggtgaaaggatatccctcgac 660

Db 601 ttatgaaaaagtacggcgttcttcaacaaaggcggcgtgaaaggatctcgtctcgac 660

Qy 661 aatgaacccgctctctgtgcgatacgcacccacgattcattcgtgtaaaaagtcgagcg 720

Db 661 aatgaacccgctctctgtgcgatacgcacgcgattcattcgtgtaaaaagtcgagcg 720

Qy 721 aaagagttggtagaccggtcagtcagtttatccaaagctgtgaaagcgtgattgacgcgggg 780

Db 721 aaagagttggtagaccggtcggtaagtttatccaaagcgcgtgaaagcgtgattgacgcgggt 780

Qy 781 gcagaggttttggccggttcttccagatttggcgctataaagatcttcaaaactgca 840

Db 781 gcggaaattttggccggttcttccagatttggcgctataaagatcttcaaaactgca 840

Qy 841 cctgattggactctgtaaaaggcaattatagctgttctgtagactattaccttgatcaa 900

Db 841 cctgattgaaactctgtaaaaggcaactacagctggttctgtagactattaccttgatcaa 900

Qy 901 atgccccttagctcgcaagtcgaaggcaagagatgctggatgatttattgacgtcaactgg 960

Db 901 atgccccttagctcgcaagtcgaaggcaagagatgctggatgatttattgacgtcaactgg 960

Qy 961 tatccgaagcagatggcgaggcgatcagaaattacgaatgaggtagggaacgaacacg 1020

Db 961 tatccgaagcagatggcgaggcgatcagaaattacgaatgaggtagggaacgaacacg 1020

Qy 1021 aagaagccagaatgcaggcacctcgcaccttggggacccgacctataaggaagatagt 1080

Db 1021 aagaagccagaatgcaggcacctcgcaccttggggacccgacctataaggaagatagt 1080

QY 1081 tggatcgctcaatggaaacagcaggtttttgcccatactacctcgattgaagcagtcggtg 1140
DB 1081 tggatcgctcaatggaaacagcaggtttttgcccatactacctcgattgaagcagtcggtg 1140
QY 1141 gataatatatccgggaacacagctggcgaatgacogagtagctatgctatggcgcaaaat 1200
DB 1141 gataatatatccgggaacacagctggcgaatgacogagtagctatgctatggcgcaaaat 1200
QY 1201 gatattccggcggttgcgatgacccgagtgctggtgtatttttggcgaataatgatgtt 1260
DB 1201 gatattccggcggttgcgatgacccgagtgctggtgtatttttggcgaataatgatgtt 1260
QY 1261 tatatggcaaaactacggaaactaaagtggtgtaacaaactagcttagtgcgcttac 1320
DB 1261 tatatggcaaaactacggaaactaaagtggtgtaacaaactagcttagtgcgcttac 1320
QY 1321 aagctttatgcgaattatgacgggaataaaactactcttctggtgataccagtggttagtgac 1380
DB 1321 aagctttatgcgaattatgacgggaataaaactactcttctggtgataccagtggttagtgac 1380
QY 1381 caaacatcggtattgtcaatagctcggtccatgcttctgtaacgaatgcatcccgacaaa 1440
DB 1381 caaacatcggtattgtcaatagctcggtccatgcttctgtaacgaatgcatcccgacaaa 1440
QY 1441 gaactgcattctgtgcatgaataaaagcatggacagcgcatctgcgcccataattgat 1500
DB 1441 gaactgcattctgtcatgaataaaagcatggacagcgcatctgcgcccataattgat 1500
QY 1501 ctctcggcggaagacttacattccggttaagtagtggtggttgcgataaaacgctcg 1560
DB 1501 ctctcggcggaagacttacattccggttaagtagtggtggttgcgataaaacgctcg 1560
QY 1561 caaattaaagaagcagcccaatccagcaaaatttcaggaacacggttttacttataccgta 1620
DB 1561 caaattaaagaagcagcccaatccagcaaaatttcaggaacacggttttacttataccgta 1620
QY 1621 ccgctttgacggcatatcacattgtgctgactactggcaatgacacgtctccagtgga 1680
DB 1621 ccgctttgacggcatatcacattgtgctgactactggcaatgacacgtctccagtgga 1680
QY 1681 ggtcctgaaagcttttaagctgaaagctgaggtggtgtagtggaagtgcttacttctcgtg 1740
DB 1681 ggtcctgaaagctttgctgaggtggtgtagtggaagtgcttacttctcgtg -tg 1740
QY 1741 gatgctccagcgaggtttagatagacagctacagcgaggaacagatgaacacgctc 1800
DB 1741 gatgctccagcgaggtttagatagacagctacagcgaggaacagatgaacacgctc 1800
QY 1735 gacgctccagcgaggtttagatagacagctacagcgaggaacagatgaacacgctc 1794
DB 1735 gacgctccagcgaggtttagatagacagctacagcgaggaacagatgaacacgctc 1794
QY 1801 ttcaactgctgtagcatcccaacttgacgaaacgctcttatacggatacactacgctgcaaac 1860
DB 1795 ttgctgctgtagcatcccaacttgctgaacagctcttatacggatacactacgctgcaaac 1854
QY 1861 ggtacttactattacaaagttaacccgcaaaacaaataaaggatcgagcgaatcccaat 1920
DB 1855 ggcacttctattattataaaataaacgcaaaacaaataaaggatcgagcgaatcccaat 1914
QY 1921 attttgaaagcgttccgaagatgctgttaaacggttcccgctgctgtatgaacccgaagaa 1980
DB 1915 gctttgaaagcgttcccgccggtgtagacggttcccggtatgctgtatgaacccgaagaa 1974
QY 1981 ggcactgtaaggggaacacatttggaatccagcgagcgagcgtactccgggtgctggtat 2040
DB 1975 ggcactgtaaggggaacacatttggaatccagtgagcgggttcccggtactggtat 2034
QY 2041 gtaacgaattccacaaatccaggggattctctgacgatgacgatccaggtcccgccgga 2100
DB 2035 gtaacgaattccacaaatccaggggattccctgacgatgacgatccaggtcccgccgga 2094
QY 2101 ggtgtgataatcttcaatccgctaccgtctctctctatgatacgaacacgcaaatctc 2160
DB 2095 ggtgtgataatcttcaatccgctaccgtctctctctatgatacgaacacgcaaatctc 2154
QY 2161 tcattaaacggcaagcgtttgagcagctgctgtttaaagaaacgctgattttaagaa 2220

DB 2155 tctttaaatggcaaaagctctggagagctggtactttggaaaacgctgatttttaagaa 2214
QY 2221 acttccggaggaaggtgctgttgaatgacggcgcaatacagatcagttttgaaacagggc 2280
DB 2215 acgtccggtggttaaggttctgttgaatgacgggggcaatacagatcgggttttgaacagggc 2274
QY 2281 tggggtggtgacgatactgactacgtacagctgagcgtccgctgacccgcaacatcat 2340
DB 2275 tggggtggtgacgatactgactacgtacagctgagcgtccgctgacccgcaacatcat 2334
QY 2341 gcggttaacccaaacgcttaccgaatccgaatcgacggttagaagcaaaagcattgatgaac 2400
DB 2335 gcggttaacgaacacgctcatcaatccgaatcgacggttagaagcaaaagcattgatgaac 2394
QY 2401 tatctggttgatcaatacaggaagaataatgctctctggtcaagagggaataaaacgaat 2460
DB 2395 tactctggttgatcaatacaggaagaataatgctctctggtcaagaggatgacccgaat 2454
QY 2461 gatggcttcaagcgaatgtaggttaaaagcggcggttgcagcgcttgcacgtgacgtgac 2520
DB 2455 gatggcttcaagcgaatgtaggttaaaagcggcggttgcagcgcttgcacgtgacgtgac 2514
QY 2521 tattcccaagcagagcggaacacggttcttagttccacagaggcagaaagggcgattgca 2580
DB 2515 tattcccaagcagagcggaacacggttcttagttccacagaggcagaaagggcgattgca 2574
QY 2581 tgggataaagcaagggggggtggttacccttgatgacggcgcaacgacccgcaagggctcg 2640
DB 2575 tgggataaagcaagggggcattggttacccttgatgacggcgcaacgacccgcaagggctcg 2634
QY 2641 atcgatacgcaggggaagaagaatggtggagaggtctctatgacgcttcaacacacattcgat 2700
DB 2635 atcgatacgcaggggaagaagaatggtggagaggtctctatgacgcttcaacacacattcgat 2694
QY 2701 atagaataatgcgagtaataccagagtcgcaagattataataattactatttcgcgacatc 2760
DB 2695 atagaataatgcgagtaataccagagtcgcaagattataataattactatttcgcgacatc 2754
QY 2761 gatgtagtgaggggcaattgaagaattgacgagtgcaaggttgcgttctcttctcgt 2820
DB 2755 gatgtagtgaggggcaattgaagaattgacgagtgcaaggttgcgttctcttctcgt 2814
QY 2821 ccttgcacgagcggaaggaagaatggtctggtggggcgcaaaagctccctgagccctgt 2880
DB 2815 ccttgcacgagcggaagggcaaatggtctggtggggcgcaaaagctccctgagccctgt 2874
QY 2881 aaaaagctgtatttttaataatgcagacgctttgacgaatgtgcacaaattgaacaaatcg 2940
DB 2875 aaaaagctgtatttttaataatgcagacgctttgacgaatgtgcacaaattgaacaaatcg 2934
QY 2941 atttgggtatggaattcttctggtccgggttgcgtatccgggagcagagtagtggtgatt 3000
DB 2935 atctgggtctggaactctgttctccgactggttccggagtagagtagtggtgatt 2994
QY 3001 ttgagcttgcgttcttccgcaacagcaggtgattacagcccgcaaaatttcaaaatacga 3060
DB 2995 ttgagcttgcgttcttccgcaacgagcagcagcagcagcagcagcagcagcagcagcag 3054
QY 3061 gacctgtgtgattgggcaaggaacaaagctag-tgcccattgagcgaataatgacccgat 3119
DB 3055 gacctgtgtacattgggcaaggaacaaagctag-tgcccattgagcgaataatgacccgat 3114
QY 3120 cccggacccctgatttgatgaaggcgatcaagctcattggagcgtggttgcgtacatgga 3179
DB 3115 cccggacccctgatttgatgaaggcgatcaagctcattggagcgtggttgcgtacatgga 3174
QY 3180 tgggagatttgcgagacggcgaacaaacag--ccttgagcattcgtgaaaaagtgat 3237
DB 3175 tgggagatttgcgagacggcgaacaaacag--ccttgagcattcgtgaaaaagtgat 3234
QY 3238 aatcatccgaacgctcattacgctgagctcccaacgaacttcaaacgctatggcatt 3297

XX (NOVO) NOVO NORDISK A/S.
 XX Jorgensen PL, Schulein M, Hansen C;
 XX WPI: 1991-238020/32.
 DR P-PSDB; AAR13227.
 XX Enzyme exhibiting cellulase activity from *Bacillus* sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg;
 PT fabrics.
 XX Claim 14; Page 81; 96pp; English.
 PS
 XX The DNA (SEQ ID NO: 1) was sequenced from plasmid pPL517 which was
 CC prep'd. from genomic DNA isolated from *Bacillus* spp. PL236. The
 CC sequence contains three potential initiation sites at positions
 CC 677, 737 and 749, but only the ATG at 677 is preceded by a ribo-
 CC some binding site. The DNA may be inserted into an expression
 CC vector for prodn. of recombinant endoglucanase in *E. coli* cells.
 CC The enzyme is useful as a cellulytic agent.
 CC See also AAQ13002 and AAQ13003.
 XX
 SQ Sequence 2977 BP; 775 A; 737 C; 821 G; 644 T; 0 other;
 Query Match 14.8%; Score 502.2; DB 12; Length 2977;
 Best Local Similarity 62.9%; Pred. No. 8.5e-184;
 Matches 966; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 120 aacggcaagactattactattaaagttagatatacattcaagatcgttaagcctattagccc 179
 Db 769 agctcgcgagtgtcatttccacgatttaacacgagtcggaacgtgcagcagccc 828
 QY 180 ttatatatacgtgatacatttggcggcgatgaaatattgctgcagacgact 239
 Db 829 caattattacggaaccaatcagatctgagcggcagcagacgtggtatcccgagct 888
 QY 240 tggtagaacgagatcacgatacaactgggaaacaaatattgctcaatgcaggaagtga 299
 Db 889 cggagcgaacggctgacggtttacactgggagaaacacgcatccgagcgggaaggga 948
 QY 300 ctggcagcaactcagcagataactattatcagtaaatgggtgctgcacacagccgaatg 359
 Db 949 ctggcttattacagcagatgtttctctcgcgcaacgggtgttccacacacccagctg 1008
 QY 360 tgaagacgagcagcgtgacgacttcttcatgaccaatcgtcgaagcttgccactta 419
 Db 1009 cgacaagcggggcggtgttaccgcttttcacgataaaatcttggagaatggagctta 1068
 QY 420 tcttttagttacgttgcgagtgccggttatgtggttaaggaatggaacggaagtgtgca 479
 Db 1069 ctccattgtacgtgcaaatggcgggttatgtgtcccggtgataagaacggtccagtga 1128
 QY 480 ggaagcgaagaaagccctccgctcgttggaaatcaggtcgtgaaccccaaaatgcacc 539
 Db 1129 cgaggtgagacgctccgctccacgctcgttgggaataaggtcgtgatttcccaaaatgc 1188
 QY 540 gttccaaactacagcgtgatctgaatgacaatcgggtctctatgtgagtgctccatt 599
 Db 1189 gttctcccttcagcgtgatctgacgacgacgaagtgatatgtgataagaagttaactt 1248
 QY 600 tttagtgaacagatcacgacgtcttcaacaaagcgggggtgaaaggtatgcccctgca 659
 Db 1249 cctggtcaacgctgtatgaaacgtcttcaacgtcaacgggcatcaacgctattcgtgga 1308
 QY 660 caatgaacccgctctggtcgcatacgcacccacgattcatggtgaaagcgtcgagc 719
 Db 1309 taacgagccgctgctggttctgagacgacatcccaaggtatccatcccgagcagttacaagc 1368
 QY 720 gaaagagttggtagaccggtcagtcagttattccaaagctgtgaaagcgtattgacgggg 779
 Db 1369 ggcagaaactcgtcgttaagacatcgactgtgtcaaggcgtgtgaagaacgtcagtcgca 1428

QY 780 ggagaggtttttggcccggttctttacggatttggcgctataaagatcttcaaaactgc 839
 Db 1429 tgcgaataattcgtctccttccctttacggttccgcgcatatttctctcgcagacgc 1488
 QY 840 acctgattgggactctgtaaaagcaattatagctggttcgttagactattaccctgatca 899
 Db 1489 accgattggccgagtttgcaggcaactacagctggtttatcgattactatcgatgata 1548
 QY 900 aatgcgcttagctcgaagtcgaagcgaagagatgctgagtgatttgcagctacactg 959
 Db 1549 gatgaagaatcgtcctatcagcagaacgcaaaagatgctgagtgctgagctccactg 1608
 QY 960 gtatcccaagcagctggcgaggcagcagcagcagcagcagcagcagcagcagcagcagc 1016
 Db 1609 gtatccggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1668
 QY 1017 aacgaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1076
 Db 1669 tacgcagaaggtcgcgtacaaagcgaagatcgtatggatccggttaccaggaaga 1728
 QY 1077 tagttgagctcgaatggaaacagcagcagcagcagcagcagcagcagcagcagcagc 1136
 Db 1729 cagctggatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1788
 QY 1137 ggtggataaataattatccgggaacacagcagcagcagcagcagcagcagcagcagcagc 1196
 Db 1789 gattcagacgtattatccgggtacgaagctggcgatcacagagttcagctagcggcaga 1848
 QY 1197 aatgatatttccggcggtgattgctgagcagcagcagcagcagcagcagcagcagcagc 1256
 Db 1849 caatcatttccggcggtgattgctgagcagcagcagcagcagcagcagcagcagcagc 1908
 QY 1257 tgtttatggcaaacactactggaagcagcagcagcagcagcagcagcagcagcagcagc 1316
 Db 1909 cgtttatgcgcggaatactgagcagcagcagcagcagcagcagcagcagcagcagc 1965
 QY 1317 ttacaagctttagcgaataattatgacggaaacacacttacttctcgtgatacaccagtttag 1376
 Db 1966 ttacaagctgtagcgaataactgacggcgaataaaatcggggttcggtcgtcagcaagtga 2025
 QY 1377 tggcgaacacgctgatttgcgaatgctcgttccatgcttctgttaacgaagatccaga 1436
 Db 2026 cgcgtctagctcgaacggaagacagcgtcgttatacgtctcgttatacgtgagagaga 2085
 QY 1437 caaagaactgcactcgttgcagtaataaagcagcagcagcagcagcagcagcagcagcagc 1496
 Db 2086 ttccgaactccactgctgctgctgaataaaatttcgacgacgacgacgacgacgacgac 2145
 QY 1497 tgatcttccggcggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1556
 Db 2146 ccagctgctggtgataaaacactacacatccgggagagatgggtcttcgacaaacgg 2205
 QY 1557 ctgcgaataaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1616
 Db 2206 atccgacattacggaacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2265
 QY 1617 cgtaccgcttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1651
 Db 2266 gcttctccattgctcgttaccacattgttctga 2300

RESULT 5

AAQ14842

ID AAQ14842 standard; DNA; 1854 BP.

XX AAQ14842;

AC AAQ14842;

XX 11-FEB-1992 (first entry)

DE B.lautus endol core domain/H.Insolens 43kD endoglucanase CBD fusion.
 XX cellulose; carbohydrate binding domain; fungi; cellulase;

Db 1398 cctgctgctgtaataaaaaatttcgacgataccgacgctactttccgctgctgtg 1457
 QY 1509 cgcgaagacttacatttcgcgttaaaagtatggggttcgataaaacacgctgcgaattaa 1568
 Db 1458 tgataaacctacacatccggagagtgatggggttcgaccaaaccggtacccgacattac 1517
 QY 1569 agaagcagcgcgaacacacgcaaaatttcaggaacacgcttttactatccgctacgccttt 1628
 Db 1518 ggaacaagcagctataacgaatatttaacaacaatcaattcgcgtatccgcttccatt 1577
 QY 1629 ggcgcgcatcacattgtgctga 1651
 Db 1578 gtcggcttaccacattgtctga 1600

RESULT 6
 AAZ45347
 ID AAZ45347 standard; DNA; 2445 BP.
 XX
 AC AAZ45347;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE DNA encoding a *Bacillus* sp. mannanase enzyme.
 XX
 KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture; ss.
 XX
 OS *Bacillus* sp.
 XX
 FH Key
 FT CDS
 FT Location/Qualifiers
 FT 1..2445
 FT /*tag= a
 FT /product= "mannanase"
 FT /note= "not termination codon given"
 FT 1..87
 FT sig_peptide
 FT /*tag= b
 FT
 FN WO964619-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 10-JUN-1999; 99WO-DK00314.
 XX
 PR 10-JUN-1998; 98US-0111256.
 PR 20-OCT-1998; 98DK-0001340.
 PR 20-OCT-1998; 98DK-0001341.
 PR 28-OCT-1998; 98US-0105970.
 PR 28-OCT-1998; 98US-0106054.
 PR 23-DEC-1998; 98DK-0001725.
 PR 05-MAR-1999; 99DK-0000306.
 PR 05-MAR-1999; 99DK-0000307.
 PR 05-MAR-1999; 99DK-0000308.
 PR 05-MAR-1999; 99DK-0000309.
 PR 09-MAR-1999; 99US-0123543.
 PR 10-MAR-1999; 99US-0123623.
 PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
 XX WPI; 2000-105891/09.
 DR P-PSDB; AAY54134.
 XX

PT New mannanases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions -
 XX Disclosure; Page 230-231; 242pp; English.
 PS
 XX
 CC The present sequence encodes a mannanase enzyme (also known as
 CC mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,
 CC or material containing guar or locust bean gums (thickeners), or to
 CC reduce viscosity of mannan-containing foods or feeds). The mannanases
 CC are also used to process coffee extracts (to inhibit gel formation); in
 CC cleaning compositions (for machine washing of fabrics, as hard-surface
 CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
 CC lens or body-care compositions) where they remove mannan-containing
 CC soils and prevent binding of some soils to cellulosics; and in fabric
 CC softeners. They can also be used in oil well drilling to fracture
 CC subterranean formations.
 XX
 SQ Sequence 2445 BP; 683 A; 535 C; 642 G; 585 T; 0 other;

Query Match 10.6%; Score 431.6; DB 21; Length 2445;
 Best Local Similarity 60.5%; Pred. NO. 1.7e-128;
 Matches 729; Conservative 0; Mismatches 469; Indels 6; Gaps 1;

QY 2105 tgtacaattctacaatcgcgtactcgttctctctcatgatgacaaacgacacatttctcat 2164
 Db 680 tatataagtaactcgttgctctatgcagctccttatggcgacaaacaaacattcgcag 739
 QY 2165 taaacggcaagcggttgccgaactgctgtcttaagaaaaacgctgattttaagaaactt 2224
 Db 740 tgaatggacagggtaccgtcaactcttgacttgaaagacagagaagctcttcgtggagtga 799
 QY 2225 cgggagggcaaggtgctgttgaaatgcagcgacgacgaatacagctatgtttgaaacaggtgg 2284
 Db 800 atgtcggcatcgtaagctcctaataaggcgaaacacactaacactccactatggtgggtgg 859
 QY 2285 gctggtacgatcagctacgctcagactggagcctgcgcg-----ctgaccgccacctc 2338
 Db 860 gatggtacaatcgtattatatacagctgtacctgtgctggttcgagtcggatcccgaaacgc 919
 QY 2339 atcggttaacaaacgcttaccacatccgaatgcagcgttagaagcaaaagcattgatga 2398
 Db 920 atcaggtcgaaaaacacactggtgaatccggacgctcactgaggcgaagagcctaatta 979
 QY 2399 actatctgttgatcaatcagggaaagaatatgctctctgttcaagaggaataaaacgaaa 2458
 Db 980 attatctctagaccagtcagggaaacaaattctatcagggtcaaacccgagtgtaagacgc 1039
 QY 2459 ttgattggcttcaagccaataggttaaaagccgaggttcagcgcgttcacctgatcg 2518
 Db 1040 ccaggtggtatccatgaacaggctgggcaaatatccctgcggttatggcagttgtttatgg 1099
 QY 2519 actattcgcgaagcagagcggaacacgctttagttccacagagcagcaaaaagcgattg 2578
 Db 1100 actacagccgctcccgctagtgcatggcgcaactggaaactcgtgttgaggagcgattg 1159
 QY 2579 catgggataagcaagggggagctgttacccttgcactggcactggaaacccagcaaggttc 2638
 Db 1160 agtgggcagagatgggtgggatacattaccctccactggtcattggcaacgcgcaaggacc 1219
 QY 2639 tgatgcatacgcagggaaagaatggtggagaggttcttatgctccgattcaaacacattcg 2698
 Db 1220 tggtaaatgtacccggcaaatgagtggtgcggtttttataccctgtgcacaaacgcttg 1279
 QY 2699 atatagaatatgcgatgaatcatccagatccgaagattataaattacttctcgcgaca 2758

Db 1280 atgtgagtagctttagaagaccgggaatctgaggatttccattgattgattagcaga 1339
 QY 2759 tcatgtgattgacgggcaattgaagaagttgcaggatgcgaaggttctctgttcc 2818
 Db 1340 tggatgtatcgcgagcaattgaagcgtgcaggcagagaaatcctgtgtatgga 1399
 QY 2819 gtcccttgcagaagcgaagaaatgttctgtgtggcgccaaaggtcctgagcctg 2878
 Db 1400 gacccttcacgagcggaagcggtgttctgtgtggcgccaaaggtcctgagcgg 1459
 QY 2879 ttaaaagctgtatatttaacgacgaccgtttgacgaatgtgcacaaattgaacatc 2938
 Db 1460 caatagagctctacaggtgatgtacgatcgttacaccaatcaaccataaacaatt 1519
 QY 2939 tgattgggtatgaattctgtctcgggattggtatccggggagacagatgtggata 2998
 Db 1520 tgatatgatgtggaattcgaagcgaaggaatgtatccggggagatgtctgtgaca 1579
 QY 2999 tttgagctttgactcttaccgcaagcaggtgattacagccgcgaatttcaaaatcgc 3058
 Db 1580 tgatcagtagcatttataatctctgcggagatttcagtcaccagcacaagtagt 1639
 QY 3059 agaacctgtgtcattgagcgaaggaacaaagctagtgtccatgagcgaataaggacga 3118
 Db 1640 agcatctaaaggaattgtacaggaagaagctgtgtgctgtgctgaaaccggcatta 1699
 QY 3119 tcccgagacctatttgatgaagcgctatcaagctcattgagctggttctgctacatgt 3178
 Db 1700 tccggatcccatcagctcagctcgttcaatgcgaactgagtggttccaccctgga 1759
 QY 3179 atggagattttgagagcggcgaacaaacagacgcttgagcatctgaaagagtgata 3238
 Db 1760 ctggagactatcagggagcgcattctcaaccctatagaacacctgcgaaggtgttc 1819
 QY 3239 atcatcgaactcattacgcttgatgagctcccaacgaactaaacgtatggcatta 3298
 Db 1820 atcatgactacgtcatcaccctggatgaattgcccggagaacctgtcccgttacggatt 1879
 QY 3299 ctga 3302
 Db 1880 ctga 1883

RESULT 7

AAK55662

ID AAK55662 standard; DNA; 6415 BP.

XX AAK55662;

AC AAK55662;

XX AAK55662;

XX 30-JUL-1999 (first entry)

DT DNA sequence encoding truncated cellulases.
 DE
 DE
 XX

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing; ss.
 XX

OS Unidentified.

XX

XX EP921188-A2.

XX

XX 09-JUN-1999.

XX

XX 15-SEP-1998; 98EP-0810919.

XX

XX 19-SEP-1997; 97US-0932571.

XX

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;

XX Gibbs MD, Morgan H, Williams DP;

XX WPI; 1999-315403/27.

XX

XX

XX

DR P-PSDB; AAY13493.

XX

PT New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim

XX

PS Disclosure; Page 24-25; 65pp; English.

XX

CC The invention relates to a recombinant cellulase active protein free of
 proteinases of native thermophilic and alkaliphilic origin, comprising
 the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 length sequences, or functional equivalents. Cel B5 extends from amino
 acid A1011 to P1424 or N1426, and Cel B4/5 extends from amino
 acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 from amino acid Y39 to D481. Cel E1/2 extends from Y39 to G635, Cel
 E1/2/3 extends from Y39 to G812. Cel B6 extends from amino acid V1233 to
 K1751 and the stability region extends from amino acid E482 to G635 in
 the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 enzymes are useful in laundry detergent compositions to prevent or
 remove staining, backstaining or graying, for use on cellulosic
 materials including cotton-containing fabrics. They are especially useful
 for preventing redeposition of colorant during stonewashing, and for
 processing of textiles where cellulose breakdown is required. The new
 truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.
 XX

SQ Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 36 other;

Query Match 10.0%; Score 405.2; DB 20; Length 6415;

Best Local Similarity 55.4%; Pred. No. 1.2e-119;

Matches 897; Conservative 1; Mismatches 669; Indels 51; Gaps 4;

QY 132 tattactattaaagtagatcattcaaggatcgtaagcctattagcccttatataacgg 191

Db 4328 tgttaaatatcgatcgatcgctccggggaagaaacaaagataagcccgatatatttg 4387

QY 192 tacaatcaggatttgcaggcgatgaaatattggtccagacgacttgggtgcaaccg 251

Db 4388 agcaatcaggatataccagggtgtg---ttcacctgcagacgacttgggtggaacag 4444

QY 252 aatgacgggatacaactgggaaacaaatattgccaatgcagggaagtgaactggcagaatc 311

Db 4445 attgacgggttacaattgggagaacaaatattgccaatgcagggaagtgaactgattc 4504

QY 312 tagcgaataactatttgcagtaattggtgacctgacacaaagcgaatgtgaaagccagg 371

Db 4505 aagcgtgatttatgtgttatattatgggtataacagggaatgataagaacgttccagc 4564

QY 372 agcgtgacgacttcgtttccatgaccaatcgctgaaacttggcacttattttagttac 431

Db 4565 agctgttgtaagcaaatctcacgagcagtaataaagcaaaatgcattacgacatcac 4624

QY 432 gttcccgatgcccgttattgtgtaaggtgaaacgggaagtgcaggaaagcgaataa 491

Db 4625 attacagatgtaggttattgtgcaaaaggtggaatggtacagtgagcagtgagac 4684

QY 492 ggcaccttcgctgttggaatcaggtcgtaaacgcgaacaaatgcaccttcccaactaca 551

Db 4685 agcttcgctgcggagatgggctgaggtcaagtttaaaaaagatggtgactgtcattgca 4744

QY 552 gcctgatctgaatgacaaatcggtctatgtgtgattgagttccatttttttagtgaaca 611

Db 4745 gcctgacgtgaatgataactatgtatatgtatgtgagtttataactatctgttaataa 4804

QY 612 gtacggcactgtcttcaacaaaggcggtgtgaaaggatatccctcgacaatgaaccgc 671

Db 4805 gtatggtcgatcacgtctgcacgggaattaaaggattatacttgcacacgagccgga 4864

QY 672 tctctggtcgcatagcaccacacattcatgtgaaagtcgagagcgaagaggttgggt 731

Db 4865 ctatggttactactcctccggaattccatccacagaaggttaacctgcagtgattgat 4924

QY 732 agaccggtcagtcagtttccaaagctgtgaaagcaggttagcgcggggcagaggtttt 791
 Db 4925 aaataacgtgtgagctgcgaagtaataaagacactgttagccagatcagaaatttt 4984
 QY 792 tggccggtttttacgatttggcgcctataaagatcttcaactcgcactgtattgga 851
 Db 4985 tggacctgcatgtatgttttggatatttaacattcagaggtcgcactgactgaa 5044
 QY 852 ctctgtaaagcaattatagctgttctgttagactattacctggtacaaatgcgccttag 911
 Db 5045 tcaggttaaagaaatcacagatgtttttgagctgtgaccttgagcagatgaagaaagc 5104
 QY 912 ctgcgaagtcgaagcagagattgctggtatttcacgtacactgtgtatccccaagc 971
 Db 5105 atcgatagtttggaaaragttattggtacttgactatcacatacactgttaccggagc 5164
 QY 972 gatggcgagcattacaaattacgaatgaagtaggcaaatgac-----gaaacgaa 1022
 Db 5165 gcaggttggcgtgtgcgaatatgtttgacggtgaaatagtagtcttcaagggtgtggc 5224
 QY 1023 gaaagcgaatgcagggcacctcgaccttgttgggacccgacctataag-----1071
 Db 5225 aatagcagagtcagcagcagacgactatgggtacgacataaaaccaccagaa 5284
 QY 1072 -----gaagatagttggtgcgtcctcaatgggaacagcgagttttgccc 1112
 Db 5285 aggtcagataacagcgggagaaatagctggataaaccatgtttccagagtagtctcc 5344
 QY 1113 catactactgattgaagcagtcggtgataaataattatccgggaaaccaagctggcaat 1172
 Db 5345 actgtctcccaatacaaggcagatatagacaagtattatctgttaccacactgttat 5404
 QY 1173 gaccgagtagactgtgcggcgaataatgatatttccggcggttgcgtgacgcatgt 1232
 Db 5405 aactgagttgattatgagggaggaaccatataatcgggaggaatagcttttagcagatgt 5464
 QY 1233 gctgggtatcttggcgaataatgatgtttatatgtgcaaacactactggaagctaaaggatg 1292
 Db 5465 gttaggatattcggcaagtagtgatatacatatgacgaagtg-----gggaga 5515
 QY 1293 tgtcaacaactacgttagtcgcttacaagctttatcgcaattatcagcgaataaactc 1352
 Db 5516 ttcggggagctatgcacagcggcgtacacatttattctcaactatgatgggaagggttc 5575
 QY 1353 tacttctggtatcagctgttagtcgcaaacatcggtatgttcaatagctgcgtcca 1412
 Db 5576 gagatcaggttcaacgtgtgagcgtgagacaactgagcgttgagaacatgcggtata 5635
 QY 1413 tgcctctgaagaaatgcacccgcaaaagaaactgcactctgtgtgcatgaataaaagcat 1472
 Db 5636 tgcctcaattgagggaggaagatgctgcactgtgcatattataattacacaggaatta 5695
 QY 1473 ggacagcattcgcagcccaatttgccttccgcgcaagacttacatttccggtaa 1532
 Db 5696 tgacaggaactgaagcagagataaagatgaataataaccagggtatacacacaggtggaga 5755
 QY 1533 agtatgggggttcgataaaaaacagctgcgcaaaattaaagaagcagcgcacacacaaat 1592
 Db 5756 gatatacaggtattgacagtacaagctctcagatcaggaagatggagtgctcagtaatat 5815
 QY 1593 ttcaggcaaccgttttactatccgtaccgcttccgcttccagcgtatcacattgtgctgac 1652
 Db 5816 acaaaacacacatcacctacagtagaagtctccaaactcagcgtatcacattgttttaac 5875
 QY 1653 tactggcaatgacacgtctcctcagtggaaggtcctggaagcttttaagctgaaagctgag 1710
 Db 5876 tcttcaaatgagataaagaataaaatggagacactgtcgtgatgtataaagttgag 5933
 RESULT 8
 AAD26526
 ID AAD26526 standard; DNA; 6416 BP.
 XX

AC AAD26526;
 XX 26-MAR-2002 (first entry)
 XX Active cellulase protein, celE gene.
 DE Active cellulase protein; alkalophilic; textile processing; proteinase;
 XX detergent additive; stonewashed appearance; cotton-containing denim;
 KW CelB5; thermophilic; commercial detergent; celE gene; ds.
 XX Unidentified.
 XX Key Location/Qualifiers
 FH 634..5889
 CD5 /tag= a
 FT /product= "CelE protein"
 FT misc_feature 748..2076
 FT /tag= b
 FT /product= "DNA encoding EI protein"
 FT 748..2538
 FT /tag= c
 FT /product= "DNA encoding EI/2 protein"
 XX
 PN US6294366-B1.
 XX 25-SEP-2001.
 XX 19-AUG-1998; 98US-0136574.
 XX 19-SEP-1997; 97US-0932571.
 PR
 XX (CLRN) CLARIANT FINANCE BVI LTD.
 XX Farrington GK, Anderson P, Bergquist P, Daniels R, Gibbs MD;
 PI Morgan H, Williams DP;
 XX WPI; 2002-081780/11.
 DR P-PSDB; AAEL6324.
 XX New cellulase active protein, useful in textile processing or
 PT commercial detergents, e.g. for improving the feel or appearance of
 PT cotton-containing fabrics, is stable under conditions of alkaline pH
 PT and elevated temperatures -
 XX Disclosure; Column 37-44; 61pp; English.
 PS The present invention relates to a cellulase active protein, which is
 XX substantially free of proteinases of native thermophilic and
 XX alkalophilic origin, where the cellulase active protein consists of the
 CC CelB5 amino acid sequence. The cellulase active protein is useful for
 CC treating cellulosic materials including cotton-containing fabrics, as
 CC detergent additives. The cellulase active protein is also useful for
 CC improving the feel and/or appearance of cotton-containing fabrics, for
 CC removing surface fibers from cotton-containing knits or for imparting
 CC stonewashed appearance to cotton-containing denims. The present proteins
 CC are stable under condition of alkaline pH and elevated temperatures,
 CC thus suitable for textile processing and in commercial detergents.
 CC The present sequence is celE gene.
 XX Sequence 6416 BP; 2068 A; 1082 C; 1689 G; 1541 T; 36 other;
 SQ

Query Match 10.0%; Score 405.2; DB 24; Length 6416;
 Best Local Similarity 55.4%; Pred. No. 1.2e-119;
 Matches 897; Conservative 1; Mismatches 669; Indels 51; Gaps 4;
 QY 132 tattactattaaagtagatcacattcaagagatcgtaagcctattagccctatatacgg 191
 Db 4329 tgttaaatatcgatcgatcgtccagggggaagaaagataaagcgtatattatgg 4388
 QY 192 tacaatcagatttggcagcgatgaaataatgctgcagacactgttggcaaccg 251
 Db 4389 agcaatcaggtatccagggtgttg---ttcacctgcagacgactgttgggaacag 4445

[illegible]

QY	1293	tgtaacaacactagcttagtcgcctctaccagctttatctcgcaattatgacgcaaaaaactc	1352
Db	5517	ttcggggagctatgcacagcgcggtacacacatttatctcaactatgatggaaaggttc	5576
QY	1353	tactttcgggtatcaccaagtttagtcgcgaacacatgatatgtcaatagctcggtcca	1412
Db	5577	gagatacgggttcaacgtgtgtgagcgtgagacaactgacggttgagacaatgcgggtata	5636
QY	1413	tgctttgtaacgaatgcattccgcagacaagaactgcacatctcgttgtcatgataaaaaagcat	1472
Db	5637	tgcttcaatgagggaagaagatgattcgcactgctgcatattataattaataacaggaatta	5696
QY	1473	ggacagcgcatcgacgcgcccaattgtattcttcgcggcggaagacttacatttcgggtaa	1532
Db	5697	tgacaggaaactggaagcgagataaagatgaataataaccagggtatatacaggtgaga	5756
QY	1533	agtatgggggttcgataaaaaaacagctcgcaaatataaagaagcagcgcaatcacgcgcaaat	1592
Db	5757	gatacagcgattgacsgtacaagctctcagatcagagaagatggagatggagtgctcagtaatat	5816
QY	1593	ttcaggcaaacgcttttacttataccgtacacgcctcttgacggcatatcacattgtgctgac	1652
Db	5817	acaaacaacacataccatcagatgaagttccaaactgcaggtataccattattgttttaac	5876
QY	1653	tactggcaatgacacgtctctcagtggaaggtcctcgaaagcttttaagctgaaagctgag	1710
Db	5877	ttctcaagtagattaaagaataaaaaatggagacactgctgcatggtgtaaaagttgag	5934

RESULT 9

AAZ45348

ID AAZ45348 standard; DNA; 1488 BP.

XX AC AAZ45348;

XX DT 27-MAR-2000 (first entry)

XX DE DNA encoding a Bacillus sp. mannase enzyme.

XX KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;

XX KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;

XX KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;

XX KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;

XX KW plant material degradation; recycled waste paper; paper making pulp;

XX KW guar; locust bean gum; mannan-containing food; coffee extract;

XX KW cleaning composition; machine washing; hard-surface cleaner;

XX KW dishwashing; oral; dental; contact lens; body-care composition;

XX KW fabric softener; oil well drilling; subterranean formation fracture; ss.

OS Bacillus sp.

XX FH key Location/Qualifiers

FT CDS 1..1488

FT FT /*tag= a

FT FT /product= "mannanase"

FT FT /note= "not termination codon given"

FT FT sig_peptide 1..111

FT FT /*tag= b

XX PN WO9964619-A2.

XX PD 16-DEC-1999.

XX PF 10-JUN-1999; 99WO-DK00314.

XX PR 10-JUN-1998; 98US-0111256.

XX PR 20-OCT-1998; 98DK-0001340.

XX PR 20-OCT-1998; 98DK-0001341.

XX PR 28-OCT-1998; 98US-0105970.

XX PR 28-OCT-1998; 98US-0106054.

XX PR 23-DEC-1998; 98DK-0001725.

XX PR 05-MAR-1999; 99DK-0000306.

XX PR 05-MAR-1999; 99DK-0000307.

PR	05-MAR-1999;	99DK-0000308.
PR	05-MAR-1999;	99DK-0000309.
PR	09-MAR-1999;	99US-0123543.
PR	10-MAR-1999;	99US-0123623.
PR	10-MAR-1999;	99US-0123641.
PR	11-MAR-1999;	99US-0123642.
XX	(NOVO) NOVO-NORDISK AS.	
PA	Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;	
XX	WPI; 2000-105891/09.	
XX	P-PSDB; AAY54135.	
XX	New mannanases for treatment of textiles, plant material and coffee	
PT	extract, and in cleaning compositions -	
PT	Disclosure; Page 233-234; 242pp; English.	
XX	The present sequence encodes a mannanase enzyme (also known as	
CC	mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).	
CC	The mannanase hydrolyses galactomannans. Specifically, mannanases	
CC	hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,	
CC	glucumannans, and galactoglucomannans. The mannanase protein, or	
CC	preparations containing it, are used to improve properties of cellulose	
CC	or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based	
CC	sizes or printing pastes). They are also used to degrade or modify	
CC	plant materials (particularly recycled waste paper, paper making pulps,	
CC	or material containing guar or locust bean gums (thickeners), or to	
CC	reduce viscosity of mannan-containing foods or feeds). The mannanases	
CC	are also used to process coffee extracts (to inhibit gel formation); in	
CC	cleaning compositions (for machine washing of fabrics, as hard-surface	
CC	cleaners, for hand or machine dishwashing, also in oral, dental, contact	
CC	lens or body-care compositions) where they remove mannan-containing	
CC	soils and prevent binding of some soils to celluloses; and in fabric	
CC	softeners. They can also be used in oil well drilling to fracture	
CC	subterranean formations.	
XX	Sequence 1488 BP; 510 A; 221 C; 323 G; 434 T; 0 other;	
SQ		
	Query Match	10.0%; Score 404.2; DB 21; Length 1488;
	Best Local Similarity	58.9%; Pred. No. 9.7e-120;
	Matches 781; Conservative 0; Mismatches 518; Indels 26; Gaps 4;	
QY	1666 tatgaagccgaaaggacgcgtgaagggaaccatttgtgaatccagcggggaccggtctac	2025
Db		
Db	130 tttagcgtgaagtgcgtatttttaaatggcggtgacctataaaaattctgaaccaggtttt	189
QY	2026 tccggtgctgttatgaacaagaattccacaatccagggggttcctctgcagtgacgatt	2085
Db		
Db	190 tctgttacggatgatgtaggtcagctttgaaaaatgctctcagatgtgcgctttcaaat	249
QY	2086 cagggtcccacgccgagcgttgtaacaattctacaaatcggtcacctctctcatgatgac	2145
Db		
Db	250 gaggtctctaaagcgggtttatacaacttaaatgtgatatggcgcgattatgg----	305
QY	2146 aaacgcaccaatttctattaacacggcaaagcgtttggcgaaactgctgttaagaaaacg	2205
Db		
Db	306 aagtggaagaagttagctaagtgtatgtcaaatgagagaagctaagtactttacsatggg	365
QY	2206 gc-tgattttaagaaaacttcggaggcaaggtgtgttgaatcagcgcggaatacagt	2264
Db		
Db	366 aagtggtcttgtaaagcgtcagcaggaaggtattactaattcagggtcttaatactact	425
QY	2265 cagttttgaaacagcgtggggcgtgggtacgatatcgactacgtcagatggagcgtgcgcg	2324
Db		
Db	426 ctcgattactcttaattggacatgggtttaccattgattattatgaattacatgcacc	485
QY	2325 tgaccgccacctcatcgctgaaccacaaacgcgttaccaatccgaatggacggtagaacg	2384
Db		
Db	486 ggaaccggaaaaaccataatgtagaaagacgtttaataaccacaaatgcaacggatgaagc	545

DE DNA encoding a Humicola insolens mannanase enzyme.

XX Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture; ss.

XX Humicola insolens.

XX Key Location/Qualifiers
 CDS 1..1464
 /tag= a
 /product= "mannanase"
 /note= "not termination codon given"
 FT sig_peptide 1..63
 FT /tag= b

XX WO964619-A2.

XX 16-DEC-1999.

XX 10-JUN-1999; 99WO-DK00314.

XX 10-JUN-1998; 98US-0111256.

XX 20-OCT-1998; 98DK-0001340.

XX 20-OCT-1998; 98DK-0001341.

XX 28-OCT-1998; 98US-0105970.

XX 28-OCT-1998; 98US-0106054.

XX 23-DEC-1998; 98DK-0001725.

XX 05-MAR-1999; 99DK-0000307.

XX 05-MAR-1999; 99DK-0000308.

XX 05-MAR-1999; 99DK-0000309.

XX 09-MAR-1999; 99US-0123543.

XX 10-MAR-1999; 99US-0123623.

XX 11-MAR-1999; 99US-0123641.

XX 11-MAR-1999; 99US-0123642.

XX (NOVO) NOVO-NORDISK AS.

XX Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

XX P-PSDB; AAV54128.

XX WPI; 2000-105891/09.

XX New mannanases for treatment of textiles, plant material and coffee

XX extract, and in cleaning compositions

XX Disclosure; Page 221; 242pp; English.

XX The present sequence encodes a mannanase enzyme (also known as

XX mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).

XX The mannanase hydrolyses galactomannans. Specifically, mannanases

XX hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,

XX glucomannans, and galactoglucomannans. The mannanase protein, or

XX preparations containing it, are used to improve properties of cellulosic

XX or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based

XX sizes or printing pastes). They are also used to degrade or modify

XX plant materials (particularly recycled waste paper, paper making pulps,

XX or material containing guar or locust bean gums (thickeners), or to

XX reduce viscosity of mannan-containing foods or feeds). The mannanases

XX are also used to process coffee extracts (to inhibit gel formation); in

XX cleaning compositions (for machine washing of fabrics as hard-surface

XX cleaners, for hand or machine dishwashing, also in oral, dental, contact

XX lens or body-care compositions) where they remove mannan-containing

XX soils and prevent binding of some soils to cellulosics; and in fabric

XX softeners. They can also be used in oil well drilling to fracture

XX subterranean formations.

XX SQ Sequence 1464 Bp; 290 A; 448 C; 456 G; 270 T; 0 other;

Query Match 8.2%; Score 333.2; DB 21; Length 1464;
 Best Local Similarity 53.8%; Pred. NO. 1.1e-96;
 Matches 714; Conservative 0; Mismatches 608; Indels 6; Gaps 1;

QY 1944 gccctgtaaacggtcccgctcgctatgaagcgaagagcagctgaaggaacattgt 2003

Db 123 gccctggtcccggtcccgaccttcgaagcggaggatgcatctccatccacgaggggt 182

QY 2004 ggaatccagcgggacccggtactccggtgctgttatgaacgaattccacatccagg 2063

Db 183 tgagtcgagctcgccggtactctgttacggatagtagcgggtctcgacgagccag 242

QY 2064 ggattctgacgatgacgattcagggtccacgagcgggtgtgataatttacaatcagg 2123

Db 243 tgacaagatcacgttccacgtggacagcgagaccacacggtgtacgacctccacctccg 302

QY 2124 ctaccgttctctcatgatgacaaacgacaaatttcttataaacggcaaacggttgg 2183

Db 303 cgtggccgcatctatggcgaagcgcaccccgctgctcattaacgagcgcgcaag 362

QY 2184 cgaactgctgcttaagaaacggtgattttaagaaacttccgagagcaaggtgctgtt 2243

Db 363 tgaggtctactcccgagcggcgtgctgttcgtcgacatcgctcgccgaggtctctgt 422

QY 2244 gaatcgagcggcgaatacagatcagttttgaaacaggggtgggtgtgacatcgacta 2303

Db 423 gaaccagggcgacacaccatcgacattgtcaacaactgggagtggtacctgacgactc 482

QY 2304 cgtcagactggagcgtcgctgaccccgccacccctcctcgcgttaacaaaacgcttacc 2363

Db 483 catcacatcaccccccctcccgcccgccgaccccccacaaaataacccctcccccgtcaa 542

QY 2364 tccgaatcgacggttagaagcaaaagcattgatgaactatctgtgtatcaatacggaa 2423

Db 543 ccttgccgacgacacacgcgcggtgtgacgataccctccgctccatctacggcaa 602

QY 2424 gaatatgctctgtgtcaagaggaaataaaacgaatgatgtggttcaagccaatgtagg 2483

Db 603 gaaaatccttccgcccagcaggagcttctctggtgggaactggatcgcccaacagacgg 662

QY 2484 taaaagccggtgattgacgctgacgctgacgctgacgctgacgctgacgctgacgctgac 2543

Db 663 caaaactcccgctggtgctcgatgatgatgatgatgatgatgatgatgatgatgatgatgat 722

QY 2544 cggcttttagttccacagagcagaaaagcgtgattgcatgggataagcaagggggatcgt 2603

Db 723 aggcactgctgggtcgtcgagggagggccatcgacatcacccgctcgccgctgattgt 782

QY 2604 tacctttgcatggcactggaacacacacacacacacacacacacacacacacacacacacac 2663

Db 783 ctccggtgtgtggcactggaacacacacacacacacacacacacacacacacacacacacac 842

QY 2664 gtggagagccttctatcgctgacattacacacacacacacacacacacacacacacacacacacac 2723

Db 843 gtggagcgggtttctacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 902

QY 2724 agagtccgaagattataaattacttatttcgcgacatcgatgtgattgagggcaattgaa 2783

Db 903 gacgaatgccaaactacacgctgctgacggtgacggtgacggtgacggtgacggtgacggtgac 962

QY 2784 gaagtgcagatgcgaaggttctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2843

Db 963 gagggtgcggagcggcggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1022

QY 2844 atggttctggtggggcgcccaaggtcctgacgtgttaaaagctgtatttttaagca 2903

Db 1023 ttggttttgggtggggaggaagggcccgagggagggagggagggagggagggagggagggagggaggg 1082

QY 2904 cgaccgttttgacgaatgtgcaaaaattgaacaattgtggtggtggtggtggtggtggtggtggtggtggtggt 2963

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 13:39:59 ; Search time 261.94 Seconds
(without alignments)
3806.316 Million cell updates/sec

Title: US-09-784-554B-1
Perfect score: 4059
Sequence: 1 atgagggcgaaaaatagtag.....ccgagccgacagctaataa 4059

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	602.2	14.8	2977	2	US-07-862-588B-1
2	601.4	14.8	1854	1	US-08-361-920-28
3	601.4	14.8	1854	1	US-08-479-939-28
4	601.4	14.8	1854	1	US-08-483-432-28
5	405.2	10.0	6415	4	US-09-136-574A-2
6	93.2	2.3	11707	4	US-09-136-574A-1
7	87.8	2.2	2029	4	US-09-136-574A-46
8	69.8	1.7	1624	2	US-07-862-588B-6
9	69.8	1.7	1775	2	US-07-862-588B-5
10	50.6	1.2	7218	1	US-08-232-463-14
11	40.4	1.0	1011	3	US-09-095-163-1
12	38.2	0.9	1482	4	US-09-198-956-9
13	38.2	0.9	1482	4	US-09-198-955A-11
14	35.4	0.9	486	1	US-08-048-164A-1
15	35.4	0.9	486	1	US-08-048-164A-3
16	35.4	0.9	486	1	US-08-460-462-1
17	35.4	0.9	486	1	US-08-460-462-3
18	35.4	0.9	486	1	US-08-460-457-1
19	35.4	0.9	486	1	US-08-460-457-3
20	35.4	0.9	486	1	US-08-460-458-1
21	35.4	0.9	486	1	US-08-460-458-3
22	35.4	0.9	486	2	US-08-460-455-1
23	35.4	0.9	486	2	US-08-460-455-3
24	35.4	0.9	486	2	US-08-330-394A-1
25	35.4	0.9	486	2	US-08-330-394A-3
26	35.4	0.9	499	3	US-09-006-636-5
27	35.4	0.9	499	4	US-09-006-632-5

28	35.2	0.9	1146	4	US-09-277-716-21	Sequence 21, Appl
29	33.6	0.8	1882	1	US-08-696-349-1	Sequence 1, Appl
30	33.6	0.8	1882	5	PCT-US96-13156-1	Sequence 1, Appl
31	33	0.8	445	2	US-08-332-766A-16	Sequence 16, Appl
32	32	0.8	774	5	PCT-US91-05766-1	Sequence 1, Appl
33	32.8	0.8	2166	1	US-08-152-019A-42	Sequence 42, Appl
34	32.4	0.8	3090	4	US-08-945-983-1	Sequence 1, Appl
35	32.4	0.8	5194	2	US-08-642-846-1	Sequence 1, Appl
36	32.4	0.8	5194	4	US-09-264-604-1	Sequence 1, Appl
37	32	0.8	1811	3	US-08-482-677-7	Sequence 7, Appl
38	31.8	0.8	1502	2	US-08-506-864A-2	Sequence 2, Appl
39	31.8	0.8	1502	2	US-08-851-968-2	Sequence 2, Appl
40	31.6	0.8	2952	4	US-09-318-794A-4	Sequence 4, Appl
41	31.4	0.8	573	4	US-09-117-257-12	Sequence 12, Appl
42	31.4	0.8	573	4	US-08-945-476-12	Sequence 12, Appl
43	31.4	0.8	573	4	US-09-489-352-12	Sequence 12, Appl
44	31.4	0.8	576	4	US-09-117-257-39	Sequence 39, Appl
45	31.4	0.8	576	4	US-09-489-352-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1

US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: NO. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA: PCT/DK91/00013
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2977 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:

NAME/KEY: CDS
 LOCATION: 677..2776
 OTHER INFORMATION:
 US-07-862-588B-1

Query Match 14.8%; Score 602.2; DB 2; Length 2977;
 Best Local Similarity 62.9%; Pred. No. 4.4e-201;
 Matches 966; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 120 aacggcaagactattactattataaagtagatatacattcaagagatcgtaagcctattagccc 179
 Db 769 AGCTGCGAGTGATGTCACATTACAGGATTAATACGCAAGTCGGAACGTCAGCGATCAGGCC 828
 QY 180 ttatatatacgtgtacaatacagagattggcagcgatgaaataatagctggcagacgact 239
 Db 829 CAATATATTACGGAACCAATACAGATCTGAGCGGACGAGAACTGGTCATCCCGCAGGCT 888
 QY 240 tgggtgaacccgaatgacggatatacactgggaacaataatgtccaatgcaggagtgga 299
 Db 889 CGGAGCAACCGGCTGACCGGTTTACACTGGGAGAACACGCAATCCAGCGCGGGAAGGGA 948
 QY 300 ctggcgcaactatagcgtataactattatgcagtaataatgggtgctgacacaaagcgaatg 359
 Db 949 CTGGCTTCATTACAGGATGATTTCTCTGCGCAACGGTGGTGTCCAGACACGACTG 1008
 QY 360 tgaagaagcagagcggtagacgaactctgttcattgacaaatgctgaagcttggaactta 419
 Db 1009 CGACAAGCGGGGGGGTGTGTACCGCTTTTCACGATAAATCTTTGGAGATGGAGCTTA 1068
 QY 420 ttcttagttagcttgcgagtgccggttatgtggctgaatgagtgctgaacaaatgctgaagct 479
 Db 1069 CTCCATTGTAACGCTGCAATGGCGGTTATGTGTCGGGATAGAACGGTCCAGTTGA 1128
 QY 480 gaaagcgaagaaagcccttcgctgctgtgaatgaagtcgttaaacgcgaacaaatgcacc 539
 Db 1129 CGAGAGTGAGAGCGGCTGCTGCTCACCGGTTGGGATAAGTGCAGTTTGCACAAATGCGCC 1188
 QY 540 gtccaactacagcctgctgctgaatgacaaatcggtctatgagtgctgctgctcattt 599
 Db 1189 GTTCTCCCTTACGCTGATCTGACGACGAGCAAGTGTATATGATGAAGAAGTTAACTT 1248
 QY 600 ttagtgaaagtagcagctgcttcaacaaagcgggggtgaaagtagatgcctcga 659
 Db 1249 CTTGCTAACCGGTATGAAACGCTTCAACGTCACGGGATCAACGGTATTCGCTGGA 1308
 QY 660 caatgaaccgctctgctgctgacagcaccacgcttcattggtgaaagtcgagc 719
 Db 1309 TAACGAGCGCGGCTGTGGTCTGAGACGCACTCAAGGATTCATCCGAGGAGTTACAAGC 1368
 QY 720 gaaagagttgtagaccggtgctgctgcttccaaagcgtgaaagcgtgacgagg 779
 Db 1369 GGCAGAACTCTGCTAAGAGCATCGACTTGTCAAGCGGTCGAGAACGTCGATCGGCA 1428
 QY 780 ggcagaggtttttggccggttttttcagagtttgccctataaagatcttcaaaactgc 839
 Db 1429 TCCGGAATATPCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899
 QY 840 acctgattgggaactctgaaagcaattatagctggttcgtagactattacactgatca 899
 Db 1489 ACCGATTTGCGGATTTGCAAGCACTACAGTGTGTTTATCGATTACTATCTGGATCA 1548
 QY 900 aatgcgcttagctgcgaatgcgaagcaagagattgctggtgatttgcgacgtacactg 959
 Db 1549 GATGAAGAATGCTCATACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1608
 QY 960 gtatcccgaaagcagatggcgagagcagacagaaatgcgaat---gaggtaggcaatgacga 1016
 Db 1609 GTATCCGGAAGACAGAGGCGGAGCGGCAATCTGCTTTGGCGGGCGGCGCAATATCCA 1668
 QY 1017 aacgaagaagcagaaatgcagcagcctgcaccttgcgacccctgagacccagcattaaagaaga 1076
 Db 1669 TACGCAAGAGGCTCGGTACAGCGCAAGATCGCTATGGGATCCGGCTTACAGGAAGA 1728

RESULT 2

US-08-361-920-28

; Sequence 28, Application US/08361920

; Patent No. 5457046

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5457046 No. 5457046disk of No. 5457046th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361.920

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00124
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3435.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1854 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus/Humicola insolens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1851
US-08-361-920-28

Query Match 14.8%; Score 601.4; DB 1; Length 1854;

Best Local Similarity 63.1%; Pred. No. 5.6e-201;
Matches 961; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 132 tattactataaagtagatcacatcaagtagtgtaagcctattagcccttatatacagg 191
Db 81 TGTCACTTTACGATTAAATACGACGAGTGCAGGAGTGCAGGATCAACCAATATTACGG 140
QY 192 tacaatcagattggcagcgatgaaatattgctccagacactgtgcaacgg 251
Db 141 ACCAATCAGGATCTGACGGGACGAGAACCTGGTTCATCCCGAGGCTCGGAGGCAACG 200
QY 252 aatgcagcgatgatacaactggaaacaaatattgccaatgcaggaatgactggcagcaatc 311
Db 201 GCTGACGGGTACACTGGGAGAACAAACATCCAGCGCGGAGGACTGGCTTCATTA 260
QY 312 tagcgaactattatgagtaagtggtgctgcacaaagccgaatgtaaaagccagg 371
Db 261 CAGCGATGATTTCTCTGCGGCAACGGTGGTGTTCAGACACCGACTGCGCAACGCGGG 320
QY 372 agcgtgacgactcgtttcatgaccaatcgctgaagcttgacatttcttttagttac 431
Db 321 GCGCGTTGTTACCGCTTTTCAGATTAATCTTTGGAGATGGAGCTTACTCCATTGTAAAC 380
QY 432 gttgcgactggccggttattggttaagtaggaaacggaagtgtgcaggaaagcgaata 491
Db 381 GCTGCAAAATGGCGGTTATGTCTCCGGGATAAGAACGGTCCAGTTGACGAGAGTGAGAC 440
QY 492 ggccttcgctcgttggaatcaggtcgtaaacgcaaaatgaacgcttccaaactaca 551
Db 441 GGCTCCGTCACCGGTTGGGATAAGGTTCGAGTTTGGGATAAGGCTTCCTCCCTTCA 500
QY 552 gctgatcgaatgacaatcgggtctctgtgtagtgatgcttcctatttttagtgaacaa 611
Db 501 GCCTGATCTGACGAGGACAGTGTATATGGATGAAGAAGTTAACTTCTCGTCAACCG 560
QY 612 gtacggaactgtcttcaacaaagcggttgaaagatagccctgcacaaatgaacccgc 671
Db 561 GTATGGAACAGCTTCAACGTCACAGGGCATCAAGCGTATTCGCTGGATTAACGAGCGCGC 620
QY 672 tctctggtgcatacaccacccagcttctggtgaaagtcgagcgaaagagtttgt 731
Db 621 GCTGTGCTGACGACGATCTCAAGGATTCATCCGAGAGTGTACAAAGCGGCAAGACTCGT 680
QY 732 agaccggtcagtcagttatccaaagctgtgaaagcgattgacgcggtgggagaggtttt 791

Db 681 CGCTAAGACATCGACTTGTCAAGCGGTGAAGAACGTCGATCCGCATGCCGAATATT 740
QY 792 tggcccggttctttacggaatttggcgcctataaagatcttcaaacctgacccctgattgga 851
Db 741 CGGTCTGCTTACGGTTTTCGGGTCATATTTGCTCTGACGAGACCGGATTTGGCC 800
QY 852 ctctgtaaaagcgaattatagctggttcgttagactattacctggatcaaatgcgccttag 911
Db 801 GAGTTTGCAGGCAACTACAGCTGGTGTTCGATTACTATCTGGATCAGATGAGAAATGC 860
QY 912 ctgcgaagtcgaagcgaagagattgctggtgtatttcgcagctacactgggtatccccgaagc 971
Db 861 TCATACCCAGACGCGCAAAAGATTGCTCGATGTGCTGGAGCTCCACTGGTATCCGGAAGC 920
QY 972 gatggcgaggagcagatacgaattacgaat---gaggtaggcaatgacgaaacgaagaagc 1028
Db 921 ACAGGGCGGAGCGCAGCAATCGTCTTTGGCGGGGGGGCAATATCGATACGCAAGAGC 980
QY 1029 cagaatgcagcaccctgcacctgtgggaccgcacctaagaagagatagttggtatcgc 1088
Db 981 TCGCGTACAAGCGCAAGATCGCTATGGGATCCGGCTTACCAGAAAGACAGCTGGATCGG 1040
QY 1089 tcaatggaacagcgagtttttggccatactacctcgtattgaagcagtcggtggtataaata 1148
Db 1041 CACATGGTTTCAAGCTACTTGGCTTAATTCGGAAGCTGCAATCTTCGATTTCAGACGTA 1100
QY 1149 ttatccgggaaccaagctggcgaatgaccgagtagtagctaaggcggcgaataatatttc 1208
Db 1101 TTATCCGGGTACGAAGCTGGCGATACAGAGTTTCAGTACGGGGGAGACATCACATTTTC 1160
QY 1209 cggcggaatgcgagatgacccatgctggtgtatcttggcgaataatgatttatagc 1268
Db 1161 GGGAGGATAGCTACCCGGGAGCGCTCGGCATTTTGGAAATATYGGCGCTTTATGCCGC 1220
QY 1269 aactactggaagcgaagatggtgtcaacaactacgttagtgcgccttacaagcttta 1328
Db 1221 GAATTACTGGCAG---ACGGAGGAGCAATACCGATTATACAGCGCTCTTTACAAGCTGTA 1277
QY 1329 tcgcaattatgacggaaaaaaccttacttccgtgtagtaccaggtttagtcgcgaacatc 1388
Db 1278 TCGCAACTACGACGGCAATAAATCGGGTTCGGGTCGATCAAAAGTGGACCGCGCTACGTC 1337
QY 1389 ggatattgcaatagctcgttccatgcttctgtaacgaatgcatccgcacaaagaactgca 1448
Db 1338 CGATACGGAGAACAGCTCGGTATACGCTTCGGTAACTGACGAGGAGAAATTCGGAACCTCA 1397
QY 1449 tctcgttgcataaataaagcattgacagcagcattgcagcccaatttgcatttccgg 1508
Db 1398 CCGTATCGTCTGAATAAAAAATTTTCGACGATCCGATCCGATCAACGCTACTTTCCAGCTGTCTGG 1457
QY 1509 cgcgaagacttaccatttccggttaaagtattgggttcgataaaacacagctcgcaaatata 1568
Db 1458 TGATAAAACCTTACACATCCGGGAGATGGGCTTCGACCAACCCGATCCGACATTCAC 1517
QY 1569 agaagcagcgcacacacacgaatttcagcgaacgcttttacttaccgctacgcgcttt 1628
Db 1518 GGAACAAGCACTATACGATATTAACAACAATCAATTCAGGTATACGCTTCTCCTCAT 1577
QY 1629 gacggcatatcacatttgcgtga 1651
Db 1578 GTCGGCTTACCACATTTGTTCTGA 1600

RESULT 3

US-08-479-939-28
; Sequence 28, Application US/08479939
; Patent No. 5686593
; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup

Db 1041 CACATGGTTTTCACAGCTACTTCCCTTAAATCCGAAGCTGCAATCTTCGATTTCAGAGTA 1100
QY 1149 ttatccggaaacaaagctggcaatgaccagagatagctatgctgagcgaataatgatatttc 1208
Db 1101 TTATCCGGGTACGAAGCTGGCGATCACAGAGTTCAGCTACGGCGAGACAAATCAATTC 1160
QY 1209 cggcgagattcgatgacgagatgctggtggtatcttgggcaaaaatgatgtttatggc 1268
Db 1161 GGGAGGATAGCTACCGGGAGCGGCTCGGCATTTTGGAAAAATATGGCGTTTATGCCGC 1220
QY 1269 aaactactggaagctaaagagtggtgcaacaactagcttagtgcccttacaagcttta 1328
Db 1221 GAATTACTGGCAG---ACGGAGGACAAATACCGATTATACAGCGCTGCTTACAAAGCTGTA 1277
QY 1329 tgcgaattatgacggaagaaacttacttcttctggtgataccagtggttagtgcgcaaacatc 1388
Db 1278 TCGCAACTACGACGGCAATAAATCGGGGTTCGGCTCGCATCAAAAGTGGACGCGCTACGTC 1337
QY 1389 ggaattgtcaatgctggtccatgcttctgtaacgaatgcatcccgcaaaaagaactgca 1448
Db 1338 CGATACGGAGAACAGCTCGGTATACGCTTCGGTAACCTGACGAGAGAAATCCGAACCTCA 1397
QY 1449 tctggttgcataaataaagcagcagcgcatctgacgcccgaatttgatcttccgg 1508
Db 1398 CCTGATCTGCTGATATAAAATTCGACGATCCGATCAACGCTACTTCCAGCTGTCTGG 1457
QY 1509 cgcgaagacttacattccggtgaagtatgggggttcgataaaaacagctcgcaaatata 1568
Db 1458 TGATAAAACCTACACATCCGGGAGATATGGGCTTCGACCAAAACCGGATCCGACATTA 1517
QY 1569 agaagcagcgaatcagcaaaatttcaggcaacccggtttacttataccgtacgccttt 1628
Db 1518 GGAACAAGCAGCTAAGCAATATTAACAACAATCAATTCACGTATACGCTTCTCCATT 1577
QY 1629 gacggcatcacattgtgctga 1651
Db 1578 GTCGGCTTACCACATGTTCTGA 1600

RESULT 5

US-09-136-574A-2

; Sequence 2, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997U001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Query Match 10.0%; Score 405.2; DB 4; Length 6416;
Best Local Similarity 55.4%; Pred. No. 2.7e-131;
Matches 897; Conservative 1; Mismatches 669; Indels 51; Gaps 4;
QY 132 tattactattaaagttagatcacattcaagatcgtaagcctattagcccttatatacagg 191
Db 4329 TGTTAAATATCGATCGATACGCTCCAGGGGAAGAAACAAAGATAAGCCCGTATTTATGG 4388
QY 192 tacaatcaggatttgccagcgcatgaaaaatatggctgccagacgacttggtggaacgcg 251
Db 4389 AGCAATACGAGTATCCAGGGTGTG---TTCACTCGCAAGACGACTGGTGGGAACAG 4445
QY 252 aatgacggatcaaacactgggaaaaacaatatgtccaatgcaggaagtgcactgagcaaatc 311
Db 4446 ATTGACGGGTTCAAATGGGAGAACAAATATGTCCAATGCAGGAGTACTGGTATCATTC 4505
QY 312 tagcgataactatttatgcagtaattggtgctgcgcacacgaagccggaatggaagaccgg 371
Db 4506 AAGCGATGATTTATGTGTTTATATATGGGTATTAACAGGGAATATAAGAACGTTCCAGC 4565
QY 372 agcggtagcactctggtttccatgaccaatcgctgaagcttgccacttatcttttagttac 431
Db 4566 AGCTGTTGTAAGCAATTTTCAGGAGCAGTCAATAAGCAAAATGCATATTACGCCATCAC 4625
QY 432 gttgcgagtgcccggttatgtggttaaggatggaacggaagtgtgcaggaaaaagcgaata 491
Db 4626 ATTACAGATGGTAGGTTATGTGGCAAAAGGATGGGAATGTTACAGTACAGCGAGTCAAGAC 4685
QY 492 ggcctctccgctcggtggaatcaggtctgtaaaccccaaaatgcacggtcccaactaca 551
Db 4686 AGCTCGCTCGCGAGATGGGCTGAGGTCAGTCAAGTTTAAAAAGATGGTGCACACTGTCAATTGCA 4745
QY 552 gctgatctgaatgacaaatcggtctctatgtgagtaggtgctccatttttttagtgaaaca 611
Db 4746 GCCTGACGTGAATGATACTATGATATATATGATGATGATGATGATGATGATGATGATGATGAT 4805
QY 612 gtacggcactgcttcaacaaagcgggggtgaaagatatgcccctcgacaaatgaacccgc 671
Db 4806 GTATGGTCGATCATCGTCTGCAACCGGGAATTAAGGATATATACTTGAACAGCGAGCCGA 4865
QY 672 tctctggtgcatacgccccacgcatcattcgtgtaaaaaagtcggagcgaagagttggt 731
Db 4866 CTATGGTTTACTACTATCCCGGATTCATCCACAGAGGTAACCTGCAGTGAATTGAT 4925
QY 732 agaccggtcagtcagtttatccaaagctgtgaaagcagattgacgcgggggagaggtttt 791
Db 4926 AAAATAATCGGTGGAGCTGGCGAAAGTAATAAAGACACTTGATCCAGATGCAGAAATTTT 4985
QY 792 tggccggttctttacggatttggcgccctataaagatatccaaactgacactgattggga 851
Db 4986 TGGACCTGCATCGTATGTTTGTGGGATATTTAAACATTTGCAGGATGCACCTGACTGTGAA 5045
QY 852 ctctgtaaaaaggaattatagctggttctgtagactattacctgtgacataaaatgcgcttag 911
Db 5046 TCAGGTAAAGGAATACAGATGCTTTTGTAGCTGGTACCTTTGAGAGATGAAGAAAGC 5105

QY 912 ctcgaagtcgaagcagagattctgctgatttgcacgtacacgtggtatcccgagc 971
Db 5106 atccgagatagtttggaaagagatttggatttgcacgtacacgtggtatcccgagc 5165
QY 972 gatggcgagagcatalcagaaatagaaatgagtaggcaatgac-----gaaacgaa 1022
Db 5166 GCAGGTTGCGGTGCGCAATATGCTTTCACGGTGCAAAATAGTACTTCAAGGGATGTGC 5225
QY 1023 gaaacgagaaatgcagagcctgcaccttgtgggaccccgacctataag----- 1071
Db 5226 AATAGCGAGGATGCGAGCAGCAGACGCTATGGGATCCGACATATAAACCACCCAGAA 5285
QY 1072 -----gaagatagttgagtcgctcaatggaacagcagattttgccc 1112
Db 5286 AGGTGAGTAAACAGCGGGGAGAAATAGCTGGATTAACCAATGGTTTCAGAGATCTTCC 5345
QY 1113 cactacacgtattgaagcagctggtgataaaatattatccgggaacacagctggcaat 1172
Db 5346 ACTGCTTCCCAATATAAAGGCGAGATATAGACAAGTATTATCTGTCACCAAACTTGCTAT 5405
QY 1173 gaccagatagctatgcccgcgaaatgatatctccggcggtggtgagatgacagctgt 1232
Db 5406 AACTGAGTTTGTATTGAGGAGGAGGAGCATTATATCGGAGGAAATAGCTTTAGCAGATGT 5465
QY 1233 gctgggtatcttggcgaataatgattttatggaactactggaagctaaaggtgg 1292
Db 5466 GTTAGGGATATTCGCAAGTATGGAGTATACATGCGCAGCAAGATG-----GGGAGA 5516
QY 1293 tgcacaaactacgttagtgcgcttacaagctttatgcgaattatgcggaacaaactc 1352
Db 5517 TCGGGGAGCTATGCACAGCGCGGTACAACTTTATCTCAACTATGATGGGAAAGTTC 5576
QY 1353 tacttccgtgtatcagctgttagtgcgaacacacgttatgttcaatagctcgttcca 1412
Db 5577 GAGATACGGTTCAACGTGTGTAGCGCTGAGACAACTACGTTGAGAACATGCCGGTATA 5636
QY 1413 tgccttgaacgaatgcatccgcgaacgaactgccttgcctgctgaataaaagcat 1472
Db 5637 TCGTTCAATTTGAGGAGAGATGATTCGACTGTGCTATATATTAATTACAGGAATTA 5696
QY 1473 gacagcgcattcgcgcgcgaatttgatcttccggcgcggaagacttaccattccggtaa 1532
Db 5697 TGACAGGAAACTGAAGCGCAGAGATAAAGATGAATAATACCGGGTATACAGAGTGGAGA 5756
QY 1533 agtatggggttcgataaaacagctcgcaattaaagaagcagcgccaaatcaacgcaat 1592
Db 5757 GATATACGGATTGACAGTACAGCTCAGATCAGGAAGATGGGAGTGTCTAGTAATAT 5816
QY 1593 ttacggcaacccgttttacttataccgtaccgcttgcggtgacggtacatcattgtgtgac 1652
Db 5817 ACAAACACACACATCCATAGAGTTCCTCAATCTGCGGTATACCATATTTGTTTAAAC 5876
QY 1653 tactggcaatgacagcttccagtggaaggtcctgaaagcttgaagctgaaagctgag 1710
Db 5877 TTCTTCAAGTAGATTAAAGATAAAATGGAGACACTGCTGCTGTTGTTAAAGTTGAG 5934

RESULT 6

US-09-136-574A-1

Sequence 1, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing

Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 11707 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-136-574A-1

Query Match 2.3%; Score 93.2; DB 4; Length 11707;
Best Local Similarity 50.4%; Pred. No. 5.6e-21;
Matches 227; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

QY 3598 ggactgtctcccaatcgcacagcagatcactaaagtgaacataatacctcaatccg 3657
Db 4038 GGTTTGAAGGTACTATACAGAACAAATGAGACAGTCCGAGCAAGTCTTATAAGCCG 4097
QY 3658 caattccaaattgtaacaaaggcacaacctccatccatccagatcaacgagttgaaattcgc 3717
Db 4098 TGGTTTAAGATAGTGAATGGAGCGAGCAGCAGTGTGTATCTTACGAGGTTAAGATAAGA 4157
QY 3718 tactactacacaaatcgacggtgacccgtgagcagactttcaactgcgactatgcgacgctg 3777
Db 4158 TACTGGTACAGTGGATGGTGACAAAGCCACAGAGTGGGTATGTGACTGGGCACAGATA 4217
QY 3778 agctgctcaaatgaaacggtgaaactggttaaaatgagaagagctgcaacccggtgcgat 3837
Db 4218 GGGCAAGCAATGTGACATTCATTTTGTGAAGCTGAGCAGCGAGTGTGAGTGGAGCGGAT 4277
QY 3838 tattatttgaagtcagtttcaattcagatgcagcgtgttttagcacctggaagagcagc 3897
Db 4278 TATTACTTGGAGGTAGGATTTAGCAGTGGAGCTGGGAGTTGGACCTTGGTAAGGACACA 4337
QY 3898 ggcgatatccaaacccgtattcataaagacagactggttcgaactataacgaagtgcgat 3957
Db 4338 GGGATATACAGTAAAGTTTAAACAAGAAATGACTGGAGCAATTACATCAGGCAGACGAC 4397
QY 3958 tattcgtataaagcagcgaacacctcatcttgcgcatcatctcaaatgatttgcgtatcat 4017
Db 4398 TGGTCATGTTGCAGAGCATGACGAATTTATGGAGAGAATGCGAAGGTAACGCTGTATGTA 4457
QY 4018 aacgctgacttgtttggggaacccgagccg 4047
Db 4458 GATGCTGTCTGGTATGGGGGAGGAGCCG 4487

RESULT 7
US-09-136-574A-46
; Sequence 46, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136.574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46

Query Match 2.2%; Score 87.8; DB 4; Length 2029;
Best Local Similarity 50.4%; Pred. No. 9.8e-20;
Matches 215; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

Qy 3625 gatactaagtgacgataatacactcaatccgcaattcccaattgtaaaaggaacaa 3684
Db 37 GAGCAAGTGGCGACACAGGTTCTATAAGCCGCTGTTAAGATAGTGAATGGAGGCGC 96
Qy 3685 acctccataccgatacaagagttgaaattcgctactactacataatcgacgtgaccgt 3744
Db 97 AGCAGTGTGATCTACAGGGTTAAGATAAGATACTGGTACACAGTGGATGGTGAAG 156
Qy 3745 gacgacatttcaactgcgactgcgacgctgactgctcaagctggaacggtgaaactg 3804
Db 157 CCACAGAGTGGCGATGTGCTAGCTGGCGACAGATAGGGGCAAGCAATGTGACATCAATTT 216
Qy 3805 gtttaaatggaagaggtgcacacggtgcgattatttgaagtcagttcaattcg 3864
Db 217 GTGAAGCTTAGCAGCGGAGTGAGTGGAGCGGATATTACCTGGAGGTAGGATTTAGCAGT 276

Qy 3865 gatcgaggcgtgttagcacctggagggaagcacgacggcgatataccaaacccgattattcataag 3924
Db 277 GGAGCTGGCAGTTGCAGCCTGGTAAAGGACACAGGGGATATACAGGTAAAGTTTAAACAAG 336
Qy 3925 acagactggtcgaaactataacgaagtgacgattatctgtataaaaggcagcaaacctca 3984
Db 337 AATGACTGGAGCAATTACAAATCAGGACGAGGACTGGTTCATGGTTGCAGAGCATGACGAAT 396
Qy 3985 ttgcccgatcatcctaaagtaccttgatcataataacggcgactctgtttggggaacccgag 4044
Db 397 TATGGAGAGAAATCGGAAGGTGACCTGTATGTAGTAGTGGTCTCTGTATATGGGGCAGGAG 456
Qy 4045 ccgacag 4051
Db 457 CCGGGAG 463

RESULT 8
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 09:39:18 ; Search time 11269 Seconds
(without alignments)
4861.482 Million cell updates/sec

Title: US-09-784-554B-1
Perfect score: 4059
Sequence: 1 atgagggcgaaaaatagtag.....ccgagccgacagctaattaa 4059

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues 27472414
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_estl:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vit:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	43.2	1.1	755	10	BI695792
2	41.4	1.0	706	10	BI195104
3	41.4	1.0	1067	10	BG340262
4	41	1.0	624	12	AZ790054
5	40.6	1.0	272	9	AA318192
6	40	1.0	416	10	BF939314
7	40	1.0	453	9	AW517942
8	40	1.0	466	9	BE140314
9	40	1.0	597	9	AA808894
10	40	1.0	661	10	BE392384
11	40	1.0	678	10	BE621757
12	40	1.0	806	10	BE548173
13	40	1.0	869	9	AL562614
14	40	1.0	910	9	AL529646
15	40	1.0	933	9	AL555910
16	40	1.0	1002	10	BF791411
17	40	1.0	1015	9	AL561103

c	18	40	1.0	1022	9	AL582389
	19	40	1.0	1081	10	BM476681
	20	40	1.0	1127	10	BM451498
c	21	39.8	1.0	540	12	AQ438438
	22	39.8	1.0	548	12	AQ439979
c	23	39	1.0	660	10	BG970771
	24	39	1.0	672	12	BH039441
	25	38.6	1.0	449	10	BE516225
c	26	38.6	1.0	499	10	BM277945
c	27	38.6	1.0	603	12	AZ288253
c	28	38.6	1.0	612	9	AL656436
c	29	38.6	1.0	997	12	CNS020X8
	30	38.6	1.0	1101	12	CNS000D1
	31	38.4	0.9	496	9	AW750993
c	32	38.4	0.9	507	9	AW253916
c	33	38.2	0.9	425	9	AI386401
c	34	38.2	0.9	431	10	W82333
c	35	38.2	0.9	432	9	AA139035
c	36	38.2	0.9	469	12	AQ667820
c	37	38.2	0.9	478	9	AI006088
c	38	38.2	0.9	508	10	BE310493
c	39	38.2	0.9	509	9	AA238513
c	40	38.2	0.9	514	9	AA764126
c	41	38.2	0.9	528	9	AI528476
c	42	38.2	0.9	567	9	AA208832
c	43	38.2	0.9	575	10	BG088257
c	44	38.2	0.9	584	9	AA615855
c	45	38.2	0.9	703	10	BI110315

ALIGNMENTS

RESULT 1
BI695792/c

LOCUS 603347050F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5374428 5',
DEFINITION mRNA sequence.

ACCESSION BI695792.1 GI:15658421
VERSION BI695792
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 755)
NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Prepared by: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM1953 row: d column: 13
High quality sequence stop: 755.
Location/Qualifiers

FEATURES
source 1..755
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5374428"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr
Library constructed by Life Technologies. Investigator

Search completed: July 12, 2002, 13:39:35
Job time: 14417 sec

Email: segreg@genoscope.cns.fr, Web : www.genoscope.cns.fr.

BASE COUNT	174 a	286 c	202 g	270 t	1 others
ORIGIN					

Query Match	1.0%;	Score 40;	DB 9;	Length 933;	
Best Local Similarity	53.9%;	Pred. No. 7.7;			
Matches 82;	Conservative 0;	Mismatches 70;	Indels 0;	Gaps 0;	
1674	agtggaaagcttcgaagcctttaagctgaaagctgagctggtgatgggaagtcattt	1733			
525	AGTGGCACAGCTCGGAAGTCTCTACAAAGAAGCTGCGGCTTGCAAGGAGAAACGCCATGT	466			
1734	atccctggatgctctccagcg9gagttgttagatcacagctacagcgggcaacagatgaaaa	1793			
465	CACCTACGTTGTAGCCAAAAGAGTGTGGCGCCGCAAAAGTCCCGCCGCCAGCTGGAGTCAG	406			
1794	cggcccttcaactgcgtgtagcatccaaacttga	1825			
405	AGGTCATTTCAAGTGTGTGACTCAAGGATGA	374			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2002, 18:47:44 ; Search time 85.28 Seconds
(without alignments)
1760.926 Million cell updates/sec

Title: US-09-784-554B-2

Perfect score: 7196

Sequence: 1 MRAKNSSNLLFKSKWLPV.....PKVTLVHNGVLVWGTEPTAN 1352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7196	100.0	1352	22	AAG63962 Amino acid sequenc
2	6413	89.1	1350	22	AAG63963 Amino acid sequenc
3	3107	43.2	695	22	AAG63964 Amino acid sequenc
4	1798	25.0	700	12	AA13227 Novel endoglucanas
5	1789.5	24.9	617	12	AA15241 B.lautus endol cor
6	1524	21.2	1751	20	AA13493 Truncated cellulase
7	1407	19.6	815	21	AA154134 Amino acid sequenc
8	1376.5	19.1	496	21	AA154135 Amino acid sequenc
9	1224	17.0	488	21	AA154128 Amino acid sequenc
10	1012	14.1	903	21	AA154137 Amino acid sequenc
11	580	8.1	551	18	AAW18790 Corrected Bacillus

12	466.5	6.5	1426	20	AAV13492 Truncated cellulase
13	459.5	6.4	782	12	AA15625 Cellulase AE-1. A
14	408	5.7	616	20	AAV13494 Truncated cellulase
15	398	5.5	476	21	AAV13494 A mannase-linker
16	388.5	5.4	499	14	AAV42122 NK-1 cellulase. B
17	376.5	5.2	493	20	AAV28850 pectate lyase-link
18	376.5	5.2	493	20	AAV43218 pectate lyase CBD
19	371	5.2	531	18	AAW15238 Scaffoldin protein
20	371	5.2	1853	19	AAW43108 C. thermocellum ce
21	338	4.7	167	17	AAV50080 Cellulose binding
22	324.5	4.5	531	16	AAW01503 60 kD endoglucanas
23	317	4.4	532	12	AA13229 Endoglucanase enco
24	314	4.4	1475	11	AAV08221 Recombinant alpha
25	296.5	4.1	537	18	AAW34562 Bankia gouldi endo
26	296.5	4.1	537	19	AAW49866 Bankia gouldi endo
27	293	4.1	586	21	AAV54126 Amino acid sequenc
28	284	3.9	487	10	AAV91007 Beta-mannase of al
29	279.5	3.9	362	21	AAV44497 Bacillus subtilis
30	278	3.9	341	22	AAE05747 Clostridium cellul
31	272.5	3.8	336	21	AAV44284 B. subtilis strain
32	267.5	3.7	361	21	AAV54136 Amino acid sequenc
33	260	3.6	428	22	AAE05748 Clostridium cellul
34	256.5	3.6	419	22	AAE05749 Mannanase. Pseudo
35	251	3.5	162	20	AAW90077 C. cellulovorans C
36	250	3.5	1684	12	AA14948 Bacterial amylase
37	246	3.4	162	15	AAE63634 Cellulose binding
38	246	3.4	163	22	AAE05745 Clostridium cellul
39	246	3.4	328	22	AAE05749 Chimeric S peptide
40	245	3.4	156	20	AAW90080 C. cellulovorans C
41	245	3.4	382	20	AAV39952 Gaussia luciferase
42	240	3.3	190	22	AAE05746 Clostridium cellul
43	239	3.3	154	20	AAW90081 C. cellulovorans C
44	231	3.2	1010	22	AAE49442 Mannanase. Cellul
45	230	3.2	470	22	AAE49443 Mannanase protein

ALIGNMENTS

RESULT 1

AAG63962
ID AAG63962 standard; Protein; 1352 AA.

AC AAG63962;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of xyloglucanase enzyme.

KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
cellulosic fiber; textile scouring.

OS Paenibacillus polymyxa.

PN WO200162903-A1.

PD 30-AUG-2001.

PF 21-FEB-2001; 2001WO-DK001116.

PR 24-FEB-2000; 2000DK-0000291.

XX (NOVO) NOVOZYMES AS.

XX Schnorr K, Jorgensen PL, Schuelein M;

XX WPI: 2001-522819/57.

XX N-PSDB; AAH75059.

XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
useful for detergent compositions, and textile or cellulose fiber
processing industries

Example 3; Page 81-85; 97pp; English.

The present sequence represents a xyloglucanase of the invention. The specification describes a xyloglucanase enzyme belonging to family 44 of glycosyl hydrolases and exhibiting a relative xyloglucanase activity of at least 30% at pH 5-8. The enzyme exhibits high performance in detergent compositions and prevents binding of certain soils to the xyloglucan left in the cellulosic material. It is stable at pH 5-10 at room temperature and has a half life of more than 50 days when incubated in a full formulated liquid detergent at 30 plus degrees celsius. The enzyme is used in detergent compositions, textile industry for improving the properties of cellulosic fibers, yarn, and woven or non-woven fabrics, preferably in textile scouring process, and in cellulose fiber processing industry for ratting of fibers e.g. hemp, jute, flax, and linen.

Sequence 1352 AA;

Query Match 100.0%; Score 7196; DB 22; Length 1352;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAKNSNLLFKRSLPVPVWACTMIVGGALPAPVVGOTAKTITIKVDFKDRKPISP 60
DB 1 mraKnsnllfkrslpvpvmactmivggalpavvvgotaktitikvdfkdrkpisp 60
QY 61 YIYGTNODLAGDENMAARRLGGNRMTCYNNENNNSNAGSDWQSSDNYLCSNGLTQAE 120
DB 61 yiygtndodlagdenmaarrlgnrmtcynnenmsnagsgdwqssdnylcsnglqgae 120
QY 121 EKPAVTTSFHQSLKLTGTVSLVTLPMAGYVAKDNGSVQSEKAPARNVQVNAKNAP 180
DB 121 ekpavttsfhqslkltgtsvltpmagyvakdngsvqsekapsarnvqvnaknap 180
QY 191 FOLOPDLNDRVYDEVFHVLNKGCTASTKAGVKGYALDNEPALMSHTPRTHGEKVGA 240
DB 191 fglpdlndrvydevfhvlvnykgctastkagvkgyaldnepalmshtprthgekvga 240
QY 241 KELVDRSVLSKAVKAIADAGAEVFGVLYGFGAYKDIQTAPDWDSDVKNGYSWFVDVYLDQ 300
DB 241 kelvdrsvlskavkaidagaevfgylyfgaykdiqtapdwdsvkngyswfvdyldq 300
QY 301 MRLSSQVEGRKLLDFVDVHYWPEAMGGIRITNEVGNDETAKKARMOAPRTLMDPTKYKDS 360
DB 301 mrlssqvegkrlldvfdvhywpeamggiritnevgndeakkarmpartlmdptkykds 360
QY 361 WIAQWSEFLPILPRLKQSDVKYYPGKLAETYSYGGENDISGGIATMDVLGILGNKV 420
DB 361 wiagwseflpilprlkqsdvdkyypgkclameysyggendisggiamtdvlgilgkndv 420
QY 421 YMANYWKLKDGNNVYSAAYKLYRNYDGKNSFTGDTSVSAQTSQSDIVNSSVHASVTVNASOK 480
DB 421 ymanywklkgdgvnnvysaayklyrnydgknsftgdtsvsaqtsdivnssvhasvtnasok 480
QY 481 ELHLVWVNSKMSDAFQDLGSAKTYISGVVGFQKNSQIKKAAPIQISGNRTYTV 540
DB 481 elhlvwnskmsdafqdlgsaktyisgvvfgfknsqikeaapitqisgnrtytv 540
QY 541 PPLTAYHVLVLTGNDTSPVGPESFKLAPAGDGKVLHLSWDASSGVVGVSVORATDENG 600
DB 541 ppltayhvlvltgndtspvgpesfklapagdgkvhlhswdassgvvgsvgratdengp 600
QY 601 FTAVASNLTTSTYDTNVTNGTSTYYKYVTKATNKGSSSNILKAVPKMPVNGPARYEAE 660
DB 601 ftavasnltsytdntvntgstyykyvktaktknkssesnilkavpkmpvngparyeae 660
QY 661 GTLKGTVSSGTCYSAGYVTFNHPGDSLTMTIQAPTAGLNLIGTRSPHDDKRTNF 720
DB 661 gtlkgtvssgtycysagyvtfhnpgdsltmtiqaptaglnligrspdhdkrtnf 720
QY 721 SLNGKAFGELLKKTADFKETSGKGLLNAGANTISFETGSGWYDIDYVRLPEAADRP 780
DB 721 slngkafgellkktadfketsgkglvllnagantisfetgsgwydydyvrlpeaadrp 780

DB 721 slngkafgellkktadfketsgkglvllnagantisfetgsgwydydyvrlpeaadrp 780
QY 781 AVTKTITNPNATVEAKALNNYLVDQYKGNLSCQBEINIDWLOANVGKKPAIALDLID 840
DB 781 avtktitnpmatveakalnnylvdqygnlsgqbeineidwloanvgkkpaaialdlid 840
QY 841 YPSRAEHGLSSTEAKEAJAWKQGGIVTFFAWHNAPKGLIDTQCKEWRGFAVSTTFD 900
DB 841 ypsraehglsssteaekajawkqggivtfawhnapkgldtqckewrgfyavsttfid 900
QY 901 IEYAMNHPSESDYKLLIRIDIVIAQOLKLQDAKVPVLPFRPLHEAGKFWWAKGPEPV 960
DB 901 ieyamnhpseedyklliridivlagqlklldakvpvlfrplheagkfwwakgpepv 960
QY 961 KKLILHMDRLINVKHNLINWNSVAPDWPGEYVDILSFDSPQAGDYSPOISKYE 1020
DB 961 kklilhmhdrilnvhknnliwnvsvapdwpgedyvdlisfdspqagdyvspiskye 1020
QY 1021 DLVALGKDKKLVAMSENGPIPDLMKAYQAHWSFATWYGFVRDQKONSLEHLKKVYN 1080
DB 1021 dlvalgkdkklvamsengpipdmlmkayqahwsfatwygdfvrdgkqnslehlkkvyn 1080
QY 1081 HPNVITLDELPTNLKTYGITEQPSVPGSTLNAAGETAKVSLSWTASANAKSYEVKRSTT 1140
DB 1081 hpnvltldelptnlktygiteqpsvpgstlnaagetakvslswtasanaksyevkrstt 1140
QY 1141 ENGAFATVADSVYGGSYTDTAVTADTYYVQVAKNDAGOTLSNTASAMPKADTQPTTG 1200
DB 1141 engafatvasdvyggsytdtadttyyqvvakndagotlsntasampkadtqpttg 1200
QY 1201 LLLQYRTADTKVNDHNLNPFQIVNKGTTTSIPINELKIRYTYTIDGREQTFNCDYATIS 1260
DB 1201 lllqyrtadtkvndhnlpqqfivnkgtttsipinelkirytytidgreqtfncdyatis 1260
QY 1261 CSKLNGLKLVKMEKATGADYILEVSNFSDAGVLAFCGGSTGDIQTRHKTDSNNYNESDDY 1320
DB 1261 csklngklvmeaatgadyilevsnfnsdagvlabpgstgdiqtrihktdwsnynesddy 1320
QY 1321 SYKGTQTSFADPKVPTLYHNGVLVWGTEPTAN 1352
DB 1321 sykgqtstsfadpkvptlyhngvlgvteptan 1352

RESULT 2

AAAG63963
ID AAG63963 standard; Protein; 1350 AA.

XX AAG63963;

XX AC AAG63963;

XX DT 29-OCT-2001 (first entry)

XX DE Amino acid sequence of xyloglucanase enzyme.

XX DE xyloglucanase; family 44; glycosyl hydrolase; detergent;

XX DE cellulose fiber; textile scouring.

XX OS Paenibacillus polymyxa.

XX FH Key Location/Qualifiers

XX FT Misc-difference 1347 /note= "Glu encoded by GAACCG"

XX FT WO200162903-A1.

XX PD 30-AUG-2001.

XX PF 21-FEB-2001; 2001WO-DK00116.

XX PR 24-FEB-2000; 2000DK-0000291.

XX PA (NOVO) NOVOZYMES AS.

PI Schnorr K, Jorgensen PL, Schuelein M;

XX WPI: 2001-522819/57.
 DR N-PSDB; AAH75060.
 XX
 PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries
 XX
 PS Disclosure; Page 86-89; 97pp; English.
 XX
 CC The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for ratting of fibers e.g. hemp, jute, flax, and
 CC linen.
 XX
 SQ Sequence 1350 AA;

Query Match 89.1%; Score 6413; DB 22; Length 1350;
 Best Local Similarity 88.9%; Pred. No. 0;
 Matches 1212; Conservative 48; Mismatches 79; Indels 24; Gaps 7;

QY 1 MRKSSNLLFRSKWLPVMACTMLVGGALPAPAVVHGQAKTITIKVDTFKDKPLSP 60
 DB 1 MAKUSSSLWSKRWLPVMACTMLVGGALPAPAVVHGQAKTITIKVDTFKDKPLSP 60
 QY 61 YIYGNQDLAGDENMAARLGRNRTGYNNENNMSNAGSDWQSSDNLCSNGGLTQAC 120
 DB 61 YIYGNQELAGDENITARRLGRNRTGYNNENNMSNAGSDWQSSDNLCSNGGLTQAC 120
 QY 121 ERPGAVTTSFHDQSLKLGTYSLVTLPMAGYVAKOCSGVQSEKAPSRWQVNAKNAP 180
 DB 121 ERPGAVTTSFHDQSLKLGTYSLVTLPMAGYVAKOCSGVQSEKAPSRWQVNAKNAP 180
 QY 181 FOLQPLDNRVYVDFEFHFLVNYKGTASTKAGVKGAYALDNEPALWSTHTRHGEKGA 240
 DB 181 FOLQPLDNRVYVDFEFHFLVNYKGTASTKAGVKGAYALDNEPALWSTHTRHGEKGA 240
 QY 241 KELVDRSVLSKAVKAIDAAGAEVFGVLYFGAYKDLQAPDWSVKNYSWFVDYLLDQ 300
 DB 241 KELVDRSVLSKAVKAIDAAGAEVFGVLYFGAYKDLQAPDWSVKNYSWFVDYLLDQ 300
 QY 301 MRLSSQVEGKRLLDVDFVHWYPEAMGGGIRITNEVGNDETAKMQAPRTLWDPTYKEDS 360
 DB 301 MRLSSQVEGKRLLDVDFVHWYPEAMGGGIRITNEVGNDETAKMQAPRTLWDPTYKEDS 360
 QY 361 WTAQWSEFLPLRLKQSDVKYYPCTKLAMTEYSYGGENDISGGIAMDVLGILGKNDV 420
 DB 361 WTAQWSEFLPLRLKQSDVKYYPCTKLAMTEYSYGGENDISGGIAMDVLGILGKNDV 420
 QY 421 YMANVWKLKDGNNVYSAAYKLYRNYDGNKSTFGDTSTYSAQTSIDVNSVHASVTNASDK 480
 DB 421 YMANVWKLKDGNNVYSAAYKLYRNYDGNKSTFGDTSTYSAQTSIDVNSVHASVTNASDK 480
 QY 481 ELHLVWVNSWDSAFDAQFDLSGAKTYISGKVGWGDKNSSQIKKAPITQSGNRFYTV 540
 DB 481 ELHLVWVNSWDSAFDAQFDLSGAKTYISGKVGWGDKNSSQIKKAPITQSGNRFYTV 540
 QY 541 PPTAYHIVLTGNDSPVEGPEFSLKAEAGDGK-VHLSWDSAGSVGVQVQRTDENG 599
 DB 541 PPTAYHIVLTGNDSPVEGPEFSLKAEAGDGK-VHLSWDSAGSVGVQVQRTDENG 599
 QY 600 PPTAVASNLSTSTYDTNVTNCTSYXKYVKTAKTNKGSSESNILKAVPKMPVNGPARYAE 659
 DB 600 PPTAVASNLSTSTYDTNVTNCTSYXKYVKTAKTNKGSSESNILKAVPKMPVNGPARYAE 659

Db 598 pfaaasnlvetsytdtnvtngtsykykitaaktkgtstasnlkavprapvdpdpyeae 657
 QY 660 EGILKGTIVSSGTYGSGAGYVTFNHPGDSLTMTTQAPTAGLYNLITIGYRSHDDKRTN 719
 Db 658 dgtlkgvtvessgtygsgagytvfnhnpagdsmttmtqaptaglynitigyrshddkrtn 717
 QY 720 FSLNGKAFGLLLKKTADFRKETSQGVLLNAGANTTSFETGCGWYDIDYVRLPEAADRPP 779
 Db 718 fslngkasgelvltktadfrketsqgvllnagantfigetgwydydyvklepaadrpp 777
 QY 780 HAVTKLTNPNAVEAKALMNYLVDQYGNLMSQEEINEIDWLQANVGKKPAIALDLI 839
 Db 778 havtkltlnpnaaveakalmnylvdqygnlmsqeedmpeidwlqanvgkkpaialdl 837
 QY 840 DYSPRAEHLGSLSTEAKATAWOKGGIVTFFAHWNAPKGLIDTQGEKWRGTYADSTTF 899
 Db 838 dyspraehlgslstetakeawdkggivtfawhwnapkgldtqgkwrgrtyadsttf 897
 QY 900 DIEYAMNHPESEDYKLLIRIDIVTAGOLKKLQDAKVPVLPRLHEAEGKFWMGAKGPEP 959
 Db 898 dieyamnhpesedyklliridivtagolkkldakvpvlprlheaegkfwmgakgpep 957
 QY 960 VKKLYILMHDRLTNVHKLNNLIWVNSVAPDWPYDGYEVDILSFDSPQAGDYSPOISKY 1019
 Db 958 vkklyilmhdrltnvhklnnliwvnsvapdwpdygyeyvdlisfdspqagdydpqisky 1017
 QY 1020 EDLVALKDKKLVAMSENGPIPD-----DLMKAYQAHWSWFATWYGFVDRGKQ 1069
 Db 1018 edlvigdkkklvcherkrtyppgsdegvsplv-----ymvw-gferrrtk 1068
 QY 1070 NSLEHLKVVNHPNVIITDELPTNLKTYGITEQSPVPGSFTLNAAGETAQVSLSWTASAN 1129
 Db 1069 qslhklkvynhpvilekltptktygiteqspvpgsftlnaagetaqvslswtasan 1128
 QY 1130 AKSYEVKRSSTENGAFATVADSVYSSYDTAVTADTYTYVQVAKNDAGOTLSNTASAM 1189
 Db 1129 aasyevkrstengafatvasdyvssydtavtadttytyvqvakndagotvntasaa 1188
 QY 1190 PKADTQPTGTLLOLYTADTKVNDNLNPOFOITVNGTTSIPINELKIRYVYIDGDRE 1249
 Db 1189 pkadtqptglvlyqyrtadtnvndnlhphqilnkgtlsvplnelkiryvyyidgdre 1248
 QY 1250 QTFNCXDYATLSCSKLNGKLVKMEKATGADYLYEVSPNSDAGVLAPGGSGDITQIRHKT 1309
 Db 1249 qtfncdyavilscsklngklvkmekatgadylyevsfnsdagvlasgsgdgtqirhka 1308
 QY 1310 DWSYNESDDYSYKGTOTSPADHPKVTLYHNGVLVWGTETPTAN 1352
 Db 1309 dwsynesddysykgtsfddhtkatlyhngvlvwtetan 1350

RESULT 3
 AAG63964
 ID AAG63964 standard; Protein; 695 AA.
 XX
 AC AAG63964;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of xyloglucanase enzyme.
 XX
 KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
 KW cellulosic fiber; textile scouring.
 XX
 OS Paenibacillus pabuli.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 239..274
 FT Misc-difference 279 /note= "36 unspecified residues encoded by NNN"
 FT Misc-difference 279 /note= "unspecified residue encoded by NTG"
 XX
 PN WO200162903-A1.

Query Match 25.0%; Score 1798; DB 12; Length 700;
 Best Local Similarity 61.28; Pred. No. 2.9e-97;
 Matches 336; Conservative 76; Mismatches 133; Indels 4; Gaps 4;

QY 12 KRSLPVMVMACTMIVGALPAPAVVHQTAKTITIKVDTFKDRKPISPIYGTNQDLAQ 71
 DB 4 tqrklrfsaalavaltvmpasvna-aasdvltfintqseraaispnlygtndqslg 62

QY 72 DENMAARLGNRTGYWENNMNAGSDWQSSDNYLCSNGGLTQAECEKPGAVTTSFH 131
 DB 63 tenwsrrlgnrlgtwennassagrdwlyhsddflcnggvpdtdcdkpgavvtafh 122

QY 132 DQSLKLTGYSLVTLPMAGYVAKDNGSVQSEKAPSAARWNOVNAKNAPFQLOPLDNDNR 191
 DB 123 dkslengaysvltqmagyvsrdknpgvdesetapsrwdkvefaknaprslqplndngq 182

QY 192 VYVDFVHFLVNMKYGTASTKAGYALDNEPALWSHTPHRIHGKVGAKELVDRSYSL 251
 DB 183 vymdeevnflvnyrnaststgikaysldnepalwsethphrihpeqlqaaelvakids 242

QY 252 KAVKAIDAGAEVGGVLYGFGAYKDLQAPDWDVSKGNYSWFVYDYLQMLRSLSQVGR 311
 DB 243 kavknvdpbaeifgpalgyfaylsldapdwpslqgnyswfidyldqmknahtqngkr 302

QY 312 LLDVDFVHWPAMGGGIRIT-NEVGNDETCKARMOAPRTLWDPTTKEDSWIAQWNEFL 370
 DB 303 lldvldvhwypaeggggrivfaggagnldtqkarvqprslwdpayqedsigtwfsyl 362

QY 371 PILPRLKOSVDKYPGTKLAMTEYSYGGENDISGGIAMDVLGILGNKNDVYMANVYKLD 430
 DB 363 plipklqsigtyyppgtklaitefsgygnhisggtadadalgifgkygvyaaenywqted 422

QY 431 GYNNVVSAAKLYRNYDGNKSTFGDTSVSAQTSIDVNSVHASVTNASDKELHLVWNKS 490
 DB 423 n-tdytsaayklyrnydgnksgfsgikvdaatsdtsenssvyavtdeensehlvlnkn 481

QY 491 MDSAFDAQDLGSKATYISGKVGWGDKNSSQIKAEAPTQISGNRFTYVPLTAYHIVL 550
 DB 482 fdppinatfqlsgdktytsgrwgdqsgdteqaaatnnnngftyltllpaysayhivl 541

QY 551 TTGNDTSPV 559
 DB 542 -kadstepv 549

RESULT 5
 AAR15241
 ID AAR15241 standard; Protein: 617 AA.
 AC AAR15241;
 DT 11-FEB-1992 (first entry)

DE B.lautus endol core domain/H.insolens 43kd endoglucanase CBD fusion.
 KW cellulose; carbohydrate binding domain; fungi; cellulase;
 KW endoglucanase; EGI; fusion gene; linking B region.
 OS Bacillus lautus.
 OS Humicola insolens.
 FH Key Location/Qualifiers
 FT Peptide 1..221
 FT /label= signal
 FT /note= "from H.insolens 43kd endoglucanase"
 FT Region 22..25
 FT /note= "First 4 N-terminal residues of 43kd endoglucanase"
 FT Domain 26..541
 FT /label= Endo_1_core
 FT /note= "from B.lautus Endo 1"
 FT Region 542..544
 FT /label= linker

FT Region 545..617
 FT /label= linking_B_region_and CBD
 FT /note= "from H.insolens 43kd endoglucanase"
 XX
 PN WO9117244-A.
 PD 14-NOV-1991.
 XX
 PF 08-MAY-1991; 91WO-DK00124.
 XX
 PR 09-MAY-1990; 90DK-0001158.
 XX
 PA (NOVO) NOVO NORDISK A/S.
 XX
 PI Woldike HF, Hagen F, Hjort CM, Hastrup S;
 DR N-PSDB; AAQ14842.
 XX
 PT New fungal (hemi) cellulose degrading enzymes - for prodn. of
 liq. fuel gas and feed protein, have specified carbohydrate
 binding domain
 XX
 PS Claim 8; Fig 15; 73pp: English.
 XX
 CC This fusion protein comprises the 43kd signal peptide and the first
 4 N-terminal amino acids of the mature ca. 43kd endoglucanase from
 H.insolens fused to the core of Endo 1 produced by B.lautus NCIMB
 40250 followed by the peptide Ile-Ser-Glu fused to the 43kd linking
 C region and carbohydrate binding domain. The fusion protein represents
 one example of a cellulose-degrading enzyme of the invention.
 CC See AAQ14838-Q14841.
 XX
 SQ Sequence 617 AA;

Query Match 24.9%; Score 1789.5; DB 12; Length 617;
 Best Local Similarity 54.0%; Pred. No. 7.6e-97;
 Matches 353; Conservative 78; Mismatches 152; Indels 71; Gaps 10;

QY 27 VGGALPAPVAVHGTAKTITIKVDTFKDRKPISPIYGTNQDLADENMAARLGNRT 86
 DB 11 vvaalpvala-aadgrsdvftintseraaispnlygtndqslgtenwsrrlgnrlt 69

QY 87 GYNWENNMNAGSDWQSSDNYLCSNGGLTQAECEKPGAVTTSFHQSLKLTGYSLVTL 146
 DB 70 gynnennassagrdwlyhsddflcnggvpdtdcdkpgavvtafhdkslqngaysvltq 129

QY 147 MAGYVAKDNGSVQSEKAPSAARWNOVNAKNAPFQLOPLDNDNRVYVDFVHFLVNMKY 206
 DB 130 magyvsrdknpgvdesetapsrwdkvefaknaprslqplndngqvydeevnflvnyg 189

QY 207 TASTKAGVGYALDNEPALWSHTPHRIHGKVGAKELVDRSVSLSKAVKAIDAGAEVFGP 266
 DB 190 naststgikaysldnepalwsethphrihpeqlqaaelvakidslkalknvdpbaeifg 249

QY 267 VLYGFGAYKDLQAPDWDVSKGNYSWFVYDYLQMLRSLSQVGRKLDVDFVHWPAMG 326
 DB 250 alygfgaylsldapdwpslqgnyswfidyldqmknahtqngkrllvldvhwypaeg 309

QY 327 GGIRIT-NEVGNDETCKARMOAPRTLWDPTTKEDSWIAQWNEFLPILPRLKOSVDKYP 385
 DB 310 ggqrlvfggagnldtqkarvqprslwdpayqedsigtwfsylplipklqsigtyy 369

QY 386 GTKLAMTEYSYGGENDISGGIAMDVLGILGNKNDVYMANVYKLDGNNVVSAAKLYRN 445
 DB 370 gtlklaitefsgygnhisggtadadalgifgkygvyaaenywqtedn-tdytsaayklyrn 428

QY 446 YDGNKSTFGDTSVSAQTSIDVNSVHASVTNASDKELHLVWNKSMDSAFDAQDLGSK 505
 DB 429 ydgnksgfsgikvdaatsdtsenssvyavtdeensehlvlnknfdpinnatfqlsgdk 488

QY 506 TYISGKVGWGDKNSSQIKAEAPTQISGNRFTYVPLTAYHIVLTTGNDTSPV--EGPE 563

Db 489 tytsgrvqfdqtsditeqaaitnnnqfytllppisayhivl-kadstepwiseips 547
 QY 564 SFKLKAEAGDGKVLHSDPSSGVGYSVQRATDENGPFPTAVASNLTERSTYTDINVTNGTS 623
 Db 548 s-----stspv-----nqp-----tststststss-- 570
 QY 624 YYYKVTAKNKGSSSESNILKAVPKMPVNGPARYEAEETGLKTIVESGTYSG 677
 Db 571 -----ppvqptpsgctaerw-----agcgngqws 596

RESULT 6

ID AAY13493 standard; Protein: 1751 AA.

XX AC AAY13493;

DT 30-JUL-1999 (first entry)

DE Truncated cellulases comprising amino acid sequence.

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing.

OS Unidentified.

XX EP921188-A2.

PN 09-JUN-1999.

PD 15-SEP-1998; 98EP-0810919.

PP 19-SEP-1997; 97US-09322571.

PR (CLRN) CLARIANT FINANCE BVI LTD.

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 PI Gibbs MD, Morgan H, Williams DP;

DR WPI; 1999-315403/27.

DR N-PSDB; AAX55662.

XX New truncated cellulase proteins, useful in detergents and for
 PT producing 'stonewashed' denim

PS Claim 7; Page 37-41; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481. Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 1751 AA;

Query Match 21.2%; Score 1524; DB 20; Length 1751;
 Best Local Similarity 52.9%; Pred. No. 1.3e-80;

Matches 283; Conservative 83; Mismatches 145; Indels 24; Gaps 6;

QY 32 PAPAVHGVQTAKTITIKVDFKDKRKPISPYIYCTNODLAGDENMAARRLGCNMTGYNWE 91
 Db 1225 ptpav-----tpdvksidtsgrtkispyiyganqdgvvvh-parrlgnnrltgnwe 1278
 QY 92 NNMNAGSDQQSDNYLCSNGGLTQAECBKPGAVTTTFHDQSLKLTYSLVTLPLWAGVY 151
 Db 1279 nmnsnagsdwyhssddymcyimgitgndknvpaavyskfheqskpnaysaitlqnvgyv 1338
 QY 152 AKDGNQSVQSEKAPARNQVNVNAKNAPFQLPQDLDNRVYVDEVFHLVKNKYGTASTK 211
 Db 1339 akdngntvsesetapsrwaevkfkkgalslqpdvndvvyvmdfinylknkygrssa 1398
 QY 212 AGVKGYALDNEPALWSHTPHRIGKEVKAKELVDRSVLSKAVKATDACAEEVFGPVLYGF 271
 Db 1399 tgikgyildnepdlwftthprlhpqkvctselinksvelakviktldpdaefgpasygf 1458
 QY 272 GAYKDLQATPDMDSVKGNYSWFVDYVLDQWRLSSQVEGRKRLDVFDFVHWYPEAMGGGIRI 331
 Db 1459 vgytltlqdpdwnqvkgnhwflswylecmkkasdsfgkrlldvidihwypeaqvgvri 1518
 QY 332 TNEVGNDETKK----ARMQAPRTLMDPTTK-----EDSWIAQNMSEFLPLPRUK 377
 Db 1519 ctd-genstsrdaiaarmqaprtlwdptyktqkgqitagenswinqwfpeylpllpnlk 1577
 QY 378 QSVDKYYPGTKLAMTEYSYGGENDISGGIAMDVLGILGKNDVYVANYWKLKGVNNYVS 437
 Db 1578 adidkyyppgtklaitefdyggkdhsggaladvgifgkygvyvmaarw---gdsqsyag 1634
 QY 438 AAYKLYRNYDGKNSTFGDTSVAQTSNVDVSSVHASVTNASDKLHLVVMNKSMDSAFDA 497
 Db 1635 aayniylndydgksrygstcvsaettdvenmpvyasiegeddstvhilinnrydklka 1694
 QY 498 QFDLSGAKTYISGKVGWGFDPKNSQIKEAPITQISGNRFTYTPVPLTAYHIVLT 552
 Db 1695 elkmantrvtygtgfygtdstssqirkmvglsnlnqntitilevpnltyvhiwts 1749

RESULT 7
 AAY54134
 ID AAY54134 standard; Protein: 815 AA.
 XX AC AAY54134;
 DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a Bacillus sp. mannanase enzyme.

KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.

OS Bacillus sp.

XX Key Location/Qualifiers
 XX Peptide 1..29
 XX Domain 44..166 /note= "signal peptide"
 XX Domain 195..300 /note= "N-terminal domain of unknown function"
 XX Domain 301..625 /note= "N-terminal domain of unknown function"
 XX Domain 626..815 /note= "catalytic domain"
 XX Domain /note= "C-terminal domain of unknown function"

CC soils and prevent binding of some soils to cellulose; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.

XX	18-NOV-1997 (first entry)	
XX	Corrected Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A.	
XX		
XX	Endoglucanase; Endo 3A; formation; localised; variation;	
KW	colour density; surface; dye; fabric; family 5; cellulose;	
KW	hydrolysatoin; p-nitrophenyl-beta-1,4-cellobioside; stone wash;	
KW	blue jeans; back staining.	
XX		
OS	Bacillus lautus.	
XX		
XX	WO9709410-A1.	
XX		
XX	13-MAR-1997.	
XX		
XX	03-SEP-1996; 96WO-DK00364.	
PF		
XX		
PR	08-SEP-1995; 95DK-0000993.	
XX		
XX	(NOVO) NOVO-NORDISK AS.	
PA		
XX		
PI	Fich M, Onishi M, Schulein M, Toft AH;	
XX		
DR	WPI; 1997-192888/17.	
XX		
PT	Localised variation of colour density in the surface of a dyed	
XX	cellulosic fabric - uses cellulase compsn. able to hydrolyse	
PT	p-nitrophenyl -beta-1,4-cellobioside	
XX		
PS	Disclosure; Pages 15-17; 23pp; English.	
XX		
CC	The present sequence is the corrected version of the incorrect	
CC	Bacillus lautus (NCIMB 40250) endoglucanase Endo 3A described in	
CC	WO9110732. Endo 3 can be used in novel method of forming localised	
CC	colour density variation on the surface of a dyed cellulosic	
CC	fabric. The method comprises agitating the fabric in an aqueous	
CC	medium (pH 6.5 to 9.0) containing a family 5 cellulose,	
CC	e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta	
CC	-1,4-cellobioside, or a family 7 cellulase, and a mechanical	
CC	abrading agent or cellulose having abrading activity. Each	
CC	cellulase displays 30 % or more of its maximum activity at pH 7.	
CC	The process is useful to provide a stone washed look to blue jeans	
CC	without back staining.	
XX		
XX	Sequence 551 AA.	
SQ		

Db 266 saalmgpgasptgwtgdaqlses-----gkwvrdqirqa----- 298
 QY 1087 LDELPLNLKTYGITEOPSPVPGSFT--LNAAGETAKVSLSWTASANAASVEKRSKSTTENGAF 1145
 Db 299 -----tggsgnptapaaptlnsatagnagvsltnavsgatsytkrattsggpy 349
 QY 1146 ATVASDVYSSYTDRAVTADTYYXQVAKNDAGOTLSWTASAMPKADTQPTTGLLLOY 1205
 Db 350 tnatvgtatsytngltngtlyyyvsaasagssasasatpaggagstgnlvqy 409
 QY 1206 RTADTKVNDHNLNPOQIVNKGTTTIPINELKIRIYYIIDGQTFNCDYATLSCSKLN 1265
 Db 410 kvgdtsatdnqmkpsfniknngtvpnlsgklrlyfckdgtadmsasfwaqigasnvs 469
 QY 1266 GKLVKMEKAATGADYILEVFSFNSDAGVLAPGGSTGDIQTRIHKTDWNSYNSDDSYKGT 1325
 Db 470 aafanf--tgsntdtvylsfagsgspagsgtgdglrmyktdwsnfneandsydg 527
 QY 1326 QTSFADHPKVTLYHNGVLWGTGP 1349
 Db 528 ktayadwnrvtlhngtvlwgtgp 551

RESULT 12

AAV13492
 ID AAV13492 standard; Protein; 1426 AA.

AC AAV13492;

DT 30-JUL-1999 (first entry)

DE Truncated cellulases Cel B4/5 and Cel B5.

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing.

OS Unidentified.

PN EP921188-A2.

PD 09-JUN-1999.

PF 15-SEP-1998; 98EP-0810919.

PR 19-SEP-1997; 97US-0932571.

PA (CLRN) CLARIANT FINANCE BVI LTD.

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

DR WPI: 1999-315403/27.

DR N-PSDB; AAX55661.

PT New truncated cellulase proteins, useful in detergents and for
 PT producing 'stonewashed' denim

PS Claim 7; Page 34-37; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAV13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAV13493; Cel E3/B5 is shown in AAV13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or

CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 1426 AA;

Query Match 6.5%; Score 466.5; DB 20; Length 1426;
 Best Local Similarity 27.6%; Pred. No. 1-2e-18;
 Matches 132; Conservative 73; Mismatches 158; Indels 115; Gaps 17;

QY 943 HEASGKFWGAKGPEPVK-----KLYILMDRLTNVHKLNLINWNSV-----A 988
 Db 129 hngtpdwffrdsqgmkskdallarkqyl-----ydvvgrykgkyawdvvnneadesq 183
 QY 989 PD-----WYP--GDEYVD---ILSFDSPQA---GDYSPOISKYEDLV-ALGKDKKL 1032
 Db 184 pdgyrrstwyqicgpeylekafiwaheadpnaklfyndyteistkrdflynmvknk-- 241
 QY 1033 AMSENGPIPDPLMKAYQAHWSWFATWYGDVFRDCKNSLEHLKVKVNH--PNV-ITLDEL 1090
 Db 242 --skgvpihgimgqshinwnp-----svseleenskiklfssipgieihitel 286
 QY 1091 PTNLKTYGITEOPSPVPGSFTLNAGET-----AKVSLSWTASA 1128
 Db 287 dmslynygsneystppqdillqrqaqykdfimlrkykgivtcvtfgikddyswnss 346
 QY 1129 NAKSYEYKRSSTENGAFATVASDVYSSYTDATVADTYYXQVAKNDAGOTLSWTASA 1188
 Db 347 skrdwpl-----lffddysak-----paywsvieaagasaaspsptva 384
 QY 1189 MP-----KADTQQPT-----TGLLQYRTADTKVNDHNLNPOQIVNKGTTSI 1231
 Db 385 tptptptvtvtatptptgtptgsglkvlykknetsastgsirpwfkivnggssv 444
 QY 1232 PINELKIRYTYTIDGREQTFNCDYATLSCSKLNGKLVKMEKAATGADYILEVFSFNSDAG 1291
 Db 445 .dlsvrkirywtytdgdkpgsacvdaqigasvntfnfklssvgsgadyylevgfssag 504
 QY 1292 VLAPGSGDIQTRIHKTDWNSYNSDDSYKGTQTSFADHPKVTLYHNGVLWGTGP 1349
 Db 505 qlqpgkdagdiqvrfrknkndwsnynqadswlqsdtdygenakvtlyvdgvlwvgdep 562

RESULT 13

AAV15625

ID AAV15625 standard; Protein; 782 AA.

AC AAV15625;

DT 17-MAR-1992 (first entry)

DE Cellulase AE-1.

XX Detergents; pharmaceuticals; deinking; carboxymethylcellulose.

OS Aeromonas strain no. 212.

XX JP03251174-A.

XX 08-NOV-1991.

XX 28-FEB-1990; 90JP-0045465.

XX 28-FEB-1990; 90JP-0045465.

XX (OJIP) OJI PAPER KK.

XX WPI: 1991-373412/51.

DR N-PSDB; AAV15178.

XX Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 PT of opt. pH when carboxymethylcellulose is used as substrate.
 XX
 PS Claim 2; Fig 3; 8pp; Japanese.
 XX
 CC The sequence was deduced from the gene which was sequenced from
 CC plasmid, pAEC 1, prep. by ligating chromosomal DNA contg. the
 CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
 CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 CC methylcellulose is the substrate. The N-terminal sequence: GHADT-
 CC has been confirmed by Edman degradation. The gene can be used to
 CC produce recombinant enzyme which is used for the effective utilis-
 CC ation of biomass resources and the mfr. of pharmaceuticals and
 CC foodstuffs, and also for the detergent and deinking of waste paper.
 XX
 SQ Sequence 782 AA;

Query Match 6.4%; Score 459.5; DB 12; Length 782;
 Best Local Similarity 25.5%; Pred. No. 1.3e-18;
 Matches 170; Conservative 81; Mismatches 226; Indels 189; Gaps 27;

Qy 809 NMLSGQEIN-EIDWLOA-NVKKP-----ATAALDLID-----YSPSRAEHLSTAEKAI 859
 Db 181 nswcgsetlnvknswadagdfgypvnqaisawtlvnyerypgafpdgslnipesngv 240
 Qy 860 AWDKQGGIV---TFAWHNAKPKGLIDTQG-----KEWWRGFYADSTTFDIEYAMNH-- 907
 Db 241 p-dldevifsgtfm-----kgvmpstglashihndnwsaf--pytnidaensmrqa 291
 Qy 908 --PESEDKLLIRIDIVIAQOLKLODAKVPVLFRLPLHEAEGRWF-----WGAKGPEP 959
 Db 292 qpstnatyavarnl-----ahlarlikpykeaeaaemwgiakaearasap 339
 Qy 960 VKKIYILMHDRLTNVHKLNLINWVNSVADPWPGDEYVDILSFD----- 1004
 Db 340 -----nv-----lytsqtpdabpggdyddirtsddryaaaaelyitaya 378
 Qy 1005 -----SYPOAGDYSP---QISKYEDLVALGDKDKLVAMSENGPIPDPLM----- 1046
 Db 379 lqdenaasykaavtssphyrevsqf-dwqstattgtlilsaqndipadlqamktnvln 437
 Qy 1047 -----KAYQAHWSWFAWYWGDFVRDGGK-----QNSLEHL 1075
 Db 438 qadgmstlngegypvllpgnkyd-----wgsnslilankmillgydydlslqldyl 489
 Qy 1076 KKV-----YNHPNVIIDELPTNLKTYGITE-----QPSYPGSFTLNAAGETAK 1119
 Db 490 kamrgmdyfmngnamrlsyi-tgygyyetttdhrwawkygsgvpyppkgwlsqgnnt 548
 Qy 1120 V---SLSWTASANAKEYEVRKSTTENGAFATVADSDVYSGSYTDTAVTADTTYQVVAKN 1176
 Db 549 vindsatrtgmpaaksyapknt-----apdawvskentinnnaplvwiskyi--q 596
 Qy 1177 DAGOTLSNTASAMPKAD-----TQQPTTGLLQYRTADTKVNDNHLNLPQFI 1223
 Db 597 dhrep1ggganqnpdpdpnppepepnpnpdpdpdpdpdpdpdpdpdpdpdpdpdpdpdp 656
 Qy 1224 VNKGTTSIPINELIRYTYIDGDEQTFNCQDYATLSCSKLNGKLVKMEKAATGADYLYLE 1283
 Db 657 vnkgaavpvlseisrlyffadgndqlqyucdwamvcslngafvkmnpkgkanadyle 716
 Qy 1284 VSPNSDAGVLAPGGSTGDIQTRHKTDWSNYNESDDSYRKGTSFADHPKVTLYHNGVL 1343
 Db 717 itfkaagslqpagkrglsqtrnhagnwanlnesndysfdptktayanwervlyhngtl 776
 Qy 1344 VWGTPEP 1349
 Db 777 vfgnep 782

RESULT 14

AAV13494
 ID AAV13494 standard; Protein; 616 AA.

XX AAV13494;

XX 30-JUL-1999 (first entry)

XX Truncated cellulase Cel E3/B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 cotton-containing fabric; stonewashing.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-0810919.

XX 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

XX WPI: 1999-315403/27.

XX N-PSDB; AAX55660.

XX New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim

XX Claim 7; Page 42-43; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAV13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAV13493; Cel E3/B5 is shown in AAV13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 616 AA;

Query Match 5.7%; Score 408; DB 20; Length 616;
 Best Local Similarity 47.0%; Pred. No. 9.5e-16;
 Matches 71; Conservative 35; Mismatches 45; Indels 0; Gaps 0;

Qy 1199 TGLLQYRTADTKVNDNHLNLPQFIQVNGKTTTIPINELIRYTYIDGREGTFNCQVAT 1258

Db 3 sgkvlyknntsaatsgsirpbfkivngsssdvlsrvkirywytdvgdkpqsavcdwaq 62

Qy 1259 LSCSKLNGKLVKMEKAATGADYLYLEVSFNSDAGVLAPGSGTGTQTRHKTDWSNYNESD 1318

Db 63 igasnvtfnfvklssgvgadylyevfssgagqlpdkdtdglqrvfnknkndwsynqad 122

Qy 1319 DYSYKGTQTSFADHPKPKVTLYHNGVLVWGTPEP 1349

Db 123 dwsvlgsmtnygenakvtlyvsgvlgwqgep 153

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:20:09 ; Search time 35.21 Seconds
(without alignments)
937.898 Million cell updates/sec

Title: US-09-784-554B-2
Perfect score: 7196
Sequence: 1 MRAKSSNLLFKRSKWLPPV.....PKVTLYHGVLVWGTEPTAN 1352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6ptodata/2/iaa/5A_COMB.pap.*
- 2: /cgn2_6ptodata/2/iaa/5B_COMB.pap.*
- 3: /cgn2_6ptodata/2/iaa/6A_COMB.pap.*
- 4: /cgn2_6ptodata/2/iaa/6B_COMB.pap.*
- 5: /cgn2_6ptodata/2/iaa/PCUOS_COMB.pap.*
- 6: /cgn2_6ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1798	25.0	700	US-07-862-588B-2	Sequence 2, Appli
2	1790.5	24.9	617	US-08-361-920-29	Sequence 29, Appl
3	1790.5	24.9	617	US-08-479-939-29	Sequence 29, Appl
4	1790.5	24.9	617	US-08-483-432-29	Sequence 29, Appl
5	1524	21.2	1751	US-09-136-574A-44	Sequence 44, Appl
6	580	8.1	551	US-09-033-537A-1	Sequence 1, Appli
7	466.5	6.5	1426	US-09-136-574A-43	Sequence 43, Appl
8	408	5.7	616	US-09-136-574A-47	Sequence 47, Appl
9	376.5	5.2	493	US-09-198-956-10	Sequence 10, Appl
10	376.5	5.2	493	US-09-198-955A-12	Sequence 12, Appl
11	338	4.7	167	PCT-US95-13813-9	Sequence 9, Appli
12	324.5	4.5	531	US-07-862-588B-7	Sequence 7, Appli
13	272.5	3.8	336	US-09-095-163-2	Sequence 2, Appli
14	251	3.5	162	US-08-048-164A-2	Sequence 2, Appli
15	251	3.5	162	US-08-460-462-2	Sequence 2, Appli
16	251	3.5	162	US-08-460-457-2	Sequence 2, Appli
17	251	3.5	162	US-08-460-458-2	Sequence 2, Appli
18	251	3.5	162	US-08-460-455-2	Sequence 2, Appli
19	251	3.5	162	US-08-330-394A-2	Sequence 2, Appli
20	251	3.5	163	US-09-006-636-7	Sequence 7, Appli
21	251	3.5	163	US-09-006-632-7	Sequence 7, Appli
22	245	3.4	156	US-08-330-394A-22	Sequence 22, Appl
23	245	3.4	382	US-09-277-716-22	Sequence 22, Appl
24	239	3.3	154	US-08-330-394A-29	Sequence 29, Appl
25	217.5	3.0	2411	US-09-268-347-36	Sequence 36, Appl
26	209.5	2.9	2353	US-09-377-155-33	Sequence 33, Appl
27	209.5	2.9	2353	US-08-913-942-4	Sequence 4, Appli

28	209.5	2.9	2353	4	US-09-669-974-33	Sequence 33, Appli
29	203	2.8	2354	4	US-09-268-347-47	Sequence 47, Appl
30	187.5	2.6	2314	4	US-09-268-347-49	Sequence 49, Appl
31	185	2.6	1222	2	US-08-682-517-15	Sequence 15, Appl
32	185	2.6	1252	2	US-08-682-517-9	Sequence 9, Appli
33	181.5	2.5	1912	1	US-08-409-995-4	Sequence 4, Appli
34	181.5	2.5	1912	3	US-08-685-467-4	Sequence 4, Appli
35	180.5	2.5	1529	2	US-08-728-470-10	Sequence 10, Appli
36	180.5	2.5	1529	4	US-08-719-641-10	Sequence 10, Appli
37	178.5	2.5	1600	2	US-08-447-031A-2	Sequence 2, Appli
38	178.5	2.5	1600	2	US-08-617-697-10	Sequence 10, Appl
39	173.5	2.4	1338	2	US-08-728-470-9	Sequence 9, Appli
40	173.5	2.4	1338	4	US-08-719-641-9	Sequence 9, Appli
41	173.5	2.4	1599	2	US-08-617-697-9	Sequence 9, Appli
42	167	2.3	592	1	US-08-217-327-8	Sequence 8, Appli
43	166	2.3	866	1	US-08-386-727-8	Sequence 8, Appli
44	166	2.3	866	2	US-08-600-452A-8	Sequence 8, Appli
45	166	2.3	1002	4	US-09-268-347-24	Sequence 24, Appli

ALIGNMENTS

RESULT 1
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5916796 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 25.0%; Score 1798; DB 2; Length 700;
Best Local Similarity 61.2%; Pred. No. 4.2e-120;

; INFORMATION FOR SEQ ID NO: 29:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 617 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-483-432-29

Query Match 24.9%; Score 1790.5; DB 1; Length 617;
Best Local Similarity 54.1%; Pred. No. 1.2e-119;
Matches 354; Conservative 77; Mismatches 152; Indels 71; Gaps 10;

QY 27 VGGALPAPVVGHTAKTIIVKDTFKDRKPIPIPIYVGTNODLAGDENMAARLGGNMT 86
DB 11 VVAALPVLALA-ADGRSDVTFINTQSERAAISPIYVGTNODLGGTNSRRLLGNRLT 69
QY 87 GYNENNNAGSDWQSSDNYLCSNGLTQAECEKPGAVTTFPHDQSLKGYSLVTLTLP 146
DB 70 GYNENNASSAGRWLHYSDDLFCNGGVPTDCKPGAVVTAFAHDKSLGNGAYSIVTLQ 129
QY 147 MAGYVAKDNGSVQSEKAPSARNOVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYK 206
DB 130 MAGYVSRDKNGPVDESETAPSRWDKVEFAKNAFSLQPDNDGQVYMDVEVFNLYRYG 189
QY 207 TASTKAGVGYALDNEPALMSHTHPRIHGEKVGAKELVDRSVLSKAVKAIDAGAEVFGP 266
DB 190 NASTSTGKAYSLDNEPALMSHTHPRIHGEKVGAKELVDRSVLSKAVKAIDAGAEVFGP 249
QY 267 VLYGFGAYKDLQAPDWDVSKGYSWFVYVYLDQMLRSSQVEGRLLDVFVHVPYEMAG 326
DB 250 ALYGFAYLSLQDAPDWDVSKGYSWFVYVYLDQMLRSSQVEGRLLDVFVHVPYEMAG 309
QY 327 GGIRIT-NEVGNDETCKARMQAPRTLMDPTTKEDSWIAQWNSBFLPILRLKQSDVKYYP 385
DB 310 GGQRIVFGGAGNIDTQARVQAPRSLWDPAEQDSNIGTWFSYLPILPLKQSSIGTYYP 369
QY 386 GTKLAMTEYSYGGENDISGGIAMPVILGILGKNDVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYK 445
DB 370 GTKLAMTEYSYGGENDISGGIAMPVILGILGKNDVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYK 428
QY 446 YDGNSTFGDTSVSAOTSDIVNSVHASVTNASDKLHLVNMKSDSAFADFDLSGAK 505
DB 429 YDGNSTFGDTSVSAOTSDIVNSVHASVTNASDKLHLVNMKSDSAFADFDLSGAK 488
QY 506 TYISGKVMGFDKNSQIKAEAPITQISGNRFTVTPPLTAYHIVLTGNDTSPV--EGPE 563
DB 489 TYTSRVWGFDTGSDITEQAATINNNQFTYTLPLSAYHIVL-KADSTEPVISEIPS 547
QY 564 SFKLKAEAGDGKVKHLSWDSSGVVGVSVQRATDENGPFATAVSNLTSTSTYDNTVNGTS 623
DB 548 S-----STSSPV-----NQP-----TSTSTSTSTSS-- 570
QY 624 YYYKVTAKTKNGSSESLKAVKMPVNGPARVEAEECTLKGTIVESSGTGYSG 677
DB 571 -----PPVQPTPSGCTAERW-----AOCGGNGWSG 596

RESULT 5

US-09-136-574A-44

; Sequence 44, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 44:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1751 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 21.2%; Score 1524; DB 4; Length 1751;
Best Local Similarity 52.9%; Pred. No. 6.7e-100;
Matches 283; Conservative 83; Mismatches 145; Indels 24; Gaps 6;

QY 32 PAPAVVHGOTAKTIIVKDTFKDRKPIPIPIYVGTNODLAGDENMAARLGGNMTGYNWE 91
DB 1225 PTPAV-----TPDVKISIDTSRGRTKISPIYIGANQDQGVVH-PARRLGGNRLTYNWE 1278
QY 92 NNNNAGSDWQSSDNYLCSNGLTQAECEKPGAVTTFPHDQSLKGYSLVTLPLMACYV 151
DB 1279 NNNNAGSDWYHSSDDYCYIMGITGNDKNVPAAVVSKPEHSIKQNAISAITLQMVGY 1338
QY 152 AKDGNQSVQSEKAPSARNOVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYGTASTK 211
DB 1339 AKDGNQSVQSEKAPSARNOVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYGTASTK 1398
QY 212 AGVGYALDNEPALMSHTHPRIHGEKVGAKELVDRSVLSKAVKAIDAGAEVFGVLYGF 271
DB 1399 TGIGYILDNEDPLWFTTHPRHPKQVTCSELINKSVELAKVIKLPDPAEIEFGPASYGF 1458
QY 272 GAYKDLQAPDWDVSKGYSWFVYVYLDQMLRSSQVEGRLLDVFVHVPYEMAGGIRI 331
DB 1459 VGLYLTQAPDWNQVKGHRWFLSWYLEQMKKASDFGRLLDVLDIHWPYEAQVGVYRI 1518
QY 332 TNEVGNDETCK-----ARMQAPRTLMDPTTK-----EDSWIAQWNSBFLPILRLK 377
DB 1519 CFD-GENSTSRDVA.TARMQAPRTLMDPTTKYTKQGITAGENSWINQWFPYLPILPNIK 1577
QY 378 QSVKYVFGTKLAMTEYSYGGENDISGGIAMPVILGILGKNDVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYK 437
DB 1578 ADIDKYVFGTKLAMTEYSYGGENDISGGIAMPVILGILGKNDVYVNAKNAFQLPDLNDRVYVDFVHFLVKNYK 1634
QY 438 RAYKLYRNDGKNSFTGDTSTVSAQTSQDIVNSVHASVTNASDKLHLVNMKSDSAFAD 497
DB 1635 RAYNYLYNDGKNSFTGDTSTVSAQTSQDIVNSVHASVTNASDKLHLVNMKSDSAFAD 1694


```
;
;
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6294366e
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match          6.5%; Score 466.5; DB 4; Length 1426;
Best Local Similarity 27.6%; Pred. No. 1.6e-24;
Matches 132; Conservative 73; Mismatches 158; Indels 115; Gaps 17;

Qy 943 HEASGKWFNGAKGPEPVK-----KLYLHMDRLTNVHKLNLWVNSV-----A 988
Db 129 HNQTPDFFRDSQMLSKDALLARLKQYI-----YDVVGRYKGVYAMDVNEAIDESQ 183
Qy 989 PD-----WYP--GDEYVD---ILSFDSYPOA---GDYSPQISKYEDLV-ALGDKKLV 1032
Db 184 PDGVRRTWYQICGPEYIEKAFIWAHEADPNAKLFYNDYNTETSTKRDFFYNNVKNLK-- 241
Qy 1033 AMSENGIPDPLMKAYQAHWFATWYGFVDRGQKNSLEHLKKVYNH--PNV-ITLDEL 1090
Db 242 --SKGVPIHGIMQSHINVNP-----SVSEIENSIKLFSSIPGIEIHTEL 286
Qy 1091 PTNLKTYGITEQPSVPGSFTLNAAGET-----AKVSLSWTASA 1128
Db 287 DMSLYNGSNENYTPPDLLQROAQYKIDFTMLRKYKIGVTCVTFWGLKDDYSWLNS 346
Qy 1129 NAKSYEVKRSTTENGAFATVADSVYSGSYTDTAVTADTTYVQVAKNDAGOTLSNTASA 1188
Db 347 SKRDWPL-----LFFDDYSAK-----PAYWSVIEAAGASASPSPTVTA 384
Qy 1189 MP-----KADTQQT-----TGLLQYRTADTKVNDHNLNPOFQIVNKGTTSI 1231
Db 385 TPTPTPTVTVTATPTTPTTGTGSGGLKVLVYKNNETSASTGSRPWFKIVNGSSV 444
Qy 1232 PINELKIRYVYIDGDEQTFNCDYATPLSCSKLNGKLVKMEKAATGADYLYEVSENSDAG 1291
Db 445 DLSRVKIRYVYDGRFQSAVCDWAQIGASNTFNFVKLSGSGVSGADYLYEVGFSGAG 504
Qy 1292 VLAFGSGTGDIQTRHKTDMSYNESDDSYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
Db 505 QLQPGKADGIQVRFNKNDWSYNQADDSWLQSMTDYGENAKVTLYVDGLVWGQEP 562

RESULT 8
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; ANDERSON, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:

;
;
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 616 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match          5.7%; Score 408; DB 4; Length 616;
Best Local Similarity 47.0%; Pred. No. 6.4e-21;
Matches 71; Conservative 35; Mismatches 45; Indels 0; Gaps 0;

Qy 1199 TGLLQYRTADTKVNDHNLNPOFQIVNKGTTSIPINELKIRYVYIDGDEQTFNCDYAT 1258
Db 3 SOVKVLYKNNETSASTGSRPWFKIVNGSSVDSLRSVKIRYVYIDGDKPKQSAVCDWAQ 62
Qy 1259 LSCSKLNGKLVKMEKAATGADYLYEVSENSDAGVLAPGSGTGDIQTRHKTDMSYNESD 1318
Db 63 IGASVNTFNVKLSGSGVSGADYLYEVGFSGAGQLQPGKTDGDIQVRFNKNDWSYNQAD 122
Qy 1319 DYSYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
Db 123 DMSWLQSMTDYGENAKVTLYVDGLVWGQEP 153

RESULT 9
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schuelein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schmorl, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
; US-09-198-956-10
```

Query Match 5.2%; Score 376.5; DB 4; Length 493;
Best Local Similarity 26.2%; Pred. No. 8e-19;
Matches 106; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 991 WYPGEYVDILSFDSPQAGDYSPQISKYEDLVALGKDKKLVAMSEN-----GP 1039
DB 135 WVDHNELYHSLNVDK-----DY-----YDGLFDVKRDAEYITFSWNYVHDGKWSMLMGS 183

QY 1040 IPDPDLKAYOAHWSFA-----TWYGDVFRDQKNSLEHLKKVYNH 1081
DB 184 SDSDNYNRTITFHHNWFENLNSRVSPFRGEGHYNNFNKIIDSGINSRMGARIRIENN 243

QY 1082 PNVTILDELPTNLKTYGITQPSVPGSFTLNAAGETAKVLSWTASANAKEYEVRKSTTE 1141
DB 244 -----LFENAKDPIVSWYSSPG-----YHVSNN-----KFVNS 273

QY 1142 NGAFATVASDVYSSYTDVAVTADTYYYQVAKNDAG-----QTLNSTASAMPKADTQ 1195
DB 274 RGSMTPTSTTTPPYYSYSLDNVDN---KSIYKQAGVKGKIQRRPPTPTSPPSANT- 329

QY 1196 OPTTG-LLQVTRADTKVNDNLNPOQIVNKGTTSPINELKIRIYYTIDGDRQTFNC 1254
DB 330 -PVSGLKVEFYNSPDSNTINSINPQFKVTNTGSSAIDLKLTLYYYTVDGKQDTFC 388

QY 1255 DYATL-----SCSKLNGKLVKMEKAATGADYILEVFSNDSAGVLAPGSTGDIQTR 1305
DB 389 DHAAIIGNSYNGITSNVKGTFVKMSSSTNNADTYLEISFT--GGTLEPGAHV-QIOGR 445

QY 1306 IHKTDWSNYESDDYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
DB 446 FAKNDWSNYTQSDYSEK-SRSQFVEMDQVYATLNGVLVWGKEP 488

RESULT 10

US-09-198-955A-12
; Sequence 12, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match 5.2%; Score 376.5; DB 4; Length 493;
Best Local Similarity 26.2%; Pred. No. 8e-19;
Matches 106; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 991 WYPGEYVDILSFDSPQAGDYSPQISKYEDLVALGKDKKLVAMSEN-----GP 1039
DB 135 WVDHNELYHSLNVDK-----DY-----YDGLFDVKRDAEYITFSWNYVHDGKWSMLMGS 183

QY 1040 IPDPDLKAYOAHWSFA-----TWYGDVFRDQKNSLEHLKKVYNH 1081
DB 184 SDSDNYNRTITFHHNWFENLNSRVSPFRGEGHYNNFNKIIDSGINSRMGARIRIENN 243

QY 1082 PNVTILDELPTNLKTYGITQPSVPGSFTLNAAGETAKVLSWTASANAKEYEVRKSTTE 1141
DB 244 -----LFENAKDPIVSWYSSPG-----YHVSNN-----KFVNS 273

QY 1142 NGAFATVASDVYSSYTDVAVTADTYYYQVAKNDAG-----QTLNSTASAMPKADTQ 1195
DB 274 RGSMTPTSTTTPPYYSYSLDNVDN---KSIYKQAGVKGKIQRRPPTPTSPPSANT- 329

QY 1196 OPTTG-LLQVTRADTKVNDNLNPOQIVNKGTTSPINELKIRIYYTIDGDRQTFNC 1254
DB 330 -PVSGLKVEFYNSPDSNTINSINPQFKVTNTGSSAIDLKLTLYYYTVDGKQDTFC 388

QY 1255 DYATL-----SCSKLNGKLVKMEKAATGADYILEVFSNDSAGVLAPGSTGDIQTR 1305
DB 389 DHAAIIGNSYNGITSNVKGTFVKMSSSTNNADTYLEISFT--GGTLEPGAHV-QIOGR 445

QY 1306 IHKTDWSNYESDDYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
DB 446 FAKNDWSNYTQSDYSEK-SRSQFVEMDQVYATLNGVLVWGKEP 488

RESULT 11

PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13813
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: BAYER=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 167 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-13813-9

Query Match

Best Local Similarity 4.7%; Score 338; DB 5; Length 167;
Matches 68; Conservative 31; Mismatches 50; Indels 14; Gaps 5;

QY 1197 PTTG-LLLQYRTADTKVNDNHLNLPQFQVKNKGTTSIPINELKIRYYITDGDREOTFNC 1255
Db 4 PVSGLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYRYIVVDGKQDQTFWCD 63
QY 1256 YATL-----SCSKLNGKLVKMEKATGADYYLYEVSFNSDAGVLAPGGSTGDIQTRI 1306
Db 64 HAAIGSGSYNGITSNVKGFVKMSSSTNNADTYLEISFT--GGTLEPGAHV-QIQGRF 120
QY 1307 HKTDSNYSNEDSYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
Db 121 AKNDSNYSQNDYSFK-SXQFVENDQVYALNGLVWGKEP 162

RESULT 12

US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796

GENERAL INFORMATION:

; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch Ieln, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/07862,588B
; FILING DATE: 19920727

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DK 164/90

; FILING DATE: 19-JAN-1990

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/DK91/00013

; FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:

; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123

; TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match

Best Local Similarity 4.5%; Score 324.5; DB 2; Length 531;
Matches 132; Conservative 66; Mismatches 206; Indels 137; Gaps 18;

QY 796 KALMNYLVDOYGNKLSQOEINEIDWLQ-----ANVCKKPAIAALDLIDYSPSRAEHGLS 851
Db 76 KSSLOWMRDNNGINVFRAAMYTSEDGVTDPFSVKNKYKEAVQASIDLALY----- 125
QY 852 STEAKAIAWMDKQGGIVTFAWH--WNAKPGGLIDTQGEKWR--GFVADSTTDEIEYAMN 906
Db 126 -----VIIDWHILSDGNPNYTKAQSARFAEQEMATLYGNTPNVIYELATS 169
QY 907 HPE-----SEYKLLIRDIID-----VIAGQKKLQD-----AKVPV-----LFRPLHEAE 946
Db 170 PTECVLGRCSSEEVITAIRSIDPDGVVIVGSPQSDIHLAADNPVSHSNVWYALH--- 226
QY 947 GKWFWMCAKGPPEPVKKLYILMHDRLTNVHKLNNLNNVNSVAPDWPYDGYVDILSFD 1006
Db 227 ---FYSGTHQ-----FLRRTIVAMNKGAAIFV-----TEWG-----TSDAS 261
QY 1007 PQAGDYSPOISKYEDLVALGK-----DK-----KLIVAMSENGPIPDPLMKAYQA 1051
Db 262 GNGGPYLPQSKWEIDFLNARKISWVNSLADKVETSAALMPGASPTGAGMPN----- 314
QY 1052 HWSWFATWYCDFYRDGKQNSLEHLKVVYHPNVITLDELPTNLKTYGITEQSPVGSFTL 1111
Db 315 -----CRMKGSGSRSPASNRRQGNPTAPAAPTN-----L 345
QY 1112 NAAGETAKVSLSWTASANAKEYEVKRSVTENGAFVAVSDVYGVSSVTDVAVTADTTYVQ 1171
Db 346 SANGGNAQVSLTNWAVSGATSYTKRATTSGGPYTNVDRGVATATSTNGLTNGTYYIV 405
QY 1172 VYAKNDAGQTLSTASAMPKADTQPTTGLLLQYRTADTKVNDNHLNLPQFQVKNKGTTSI 1231
Db 406 VRASNSAGSSANSAQAATPAGSGASTGNLVQYKVGDTSATDNQMKPSFNKNNGTTPV 465
QY 1232 PTELKIRYYITDGDREOTFNCVATLSCSKLNGKLVKMEKATCAD-YLYEVGFNSDA 1290
Db 466 NLSGLKXXXXXKDGADMSCSIDWAQI--GRTNVLLAFANFTGNTDTCCELSFSC 523
QY 1291 G 1291
Db 524 G 524

RESULT 13

US-09-095-163-2

; Sequence 2, Application US/09095163

; Patent No. 6060299

GENERAL INFORMATION:

; APPLICANT: Sreekrishna, Kotikanyadanam

; APPLICANT: Johnstone, Kevin

; APPLICANT: Saunders, Charles

; APPLICANT: Bettiol, Jean-Luc

; TITLE OF INVENTION: An Enzyme Exhibiting Mannase

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 60602990 No. 6060299disk of No. 6060299th America, Inc.

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: U.S.A.

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/095,163

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2002, 19:13:04 ; Search time 59 seconds
(without alignments)
2201.912 Million cell updates/sec

Title: US-09-784-554B-2

Perfect score: 7196

Sequence: 1 MRAKNSNLLFKRSKWLPPV.....PKVTLYHNGVLVWGTEPTAN 1352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1779	24.7	700	2 B41897	cellulase (EC 3.2.
2	1552.5	21.6	606	2 H97012	hypothetical prote
3	1459.5	20.3	1331	2 A48954	mannan endo-1,4-be
4	937	13.0	1021	2 T10748	mannan endo-1,4-be
5	874	12.1	591	2 JC7657	mannan endo-1,4-be
6	454.5	6.3	1779	2 T31085	xylanase - Caldice
7	443	6.2	915	2 A43802	cellulase (EC 3.2.
8	441.5	6.1	1039	2 S02711	cellulase (EC 3.2.
9	406.5	5.6	1742	2 T17130	cellulase (EC 3.2.
10	404.5	5.6	1711	2 T31337	cellulase (EC 3.2.
11	396.5	5.5	499	2 JN0111	1,4-beta-glucanase
12	394	5.5	508	2 G69593	cellulase (EC 3.2.
13	392.5	5.5	508	2 A26874	cellulase (EC 3.2.
14	388.5	5.4	499	2 A27198	cellulase (EC 3.2.
15	371	5.2	145	2 A41897	cellulase homolog
16	371	5.2	1854	2 S36859	c1pA protein - Clo
17	366	5.1	986	2 S12021	thermoactive cellu
18	361	5.0	535	2 A41375	cellulase (EC 3.2.
19	347	4.8	879	2 A47704	endoglucanase I (E
20	329	4.6	486	2 I40548	bifunctional cellu
21	329	4.6	505	2 S39662	endoglucanase - Er
22	320.5	4.5	504	2 S54744	cellulase (EC 3.2.
23	315	4.4	1475	2 A44765	alpha-amylase (EC
24	308	4.3	1483	2 C97012	probably celluloso
25	291	4.0	513	2 A37219	mannan endo-1,4-be
26	287.5	4.0	586	2 PC6006	scaffolding protei
27	285	4.0	1203	2 S27545	pullulanase - Ther
28	279.5	3.9	362	2 H69785	mannan endo-1,4-be
29	276.5	3.9	1481	2 S28669	amyllopullulanase p

30	275.5	3.8	1162	2 T30433	scaffolding protei
31	269.5	3.7	1230	2 S47466	cellulose 1,4-beta
32	258.5	3.6	360	2 S60268	beta-mannosidase (
33	258.5	3.6	419	2 S53374	mannanase A - pseu
34	251	3.5	1848	2 A4140	cellulose-binding
35	250	3.5	1684	2 S10789	cellulase A-180 - al
36	246.5	3.4	1959	2 AG1085	hypothetical prote
37	235	3.3	1282	2 JC4393	microbial collagen
38	223	3.1	2032	2 I39917	hypothetical prote
39	210	2.9	26926	1 I38344	titin, cardiac mus
40	207	2.9	3283	2 AC1018	large repetitive p
41	206.5	2.9	1946	2 JC6032	lactocepin (EC 3.4
42	206	2.9	1993	2 AF1450	probable peptidogl
43	205	2.8	5188	2 B85547	probable RX fami
44	201	2.8	2013	2 A11489	probable peptidogl
45	200.5	2.8	1090	2 S59077	cellulose 1,4-beta

ALIGNMENTS

RESULT 1

B41897

cellulase (EC 3.2.1.4) - Bacillus lautus

N;Alternate names: endo-1,4-beta-glucanase

C;Species: Bacillus lautus

C;Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999

C;Accession: B41897; S27499

R;Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A;Title: ccla from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-

A;Reference number: B41897; MUID:92276330

A;Accession: B41897

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-700 <HAN>

A;Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663

A;Experimental source: PL236

A;Note: sequence extracted from NCBI backbone (NCBIP:104605)

C;Function:

A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A;Pathway: cellulose degradation

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 24.7%; Score 1779; DB 2; Length 700;
Best Local Similarity 60.8%; Pred. No. 3.9e-94;
Matches 334; Conservative 75; Mismatches 136; Indels 4; Gaps 4;

QY 12 KRKWLPPVMACTMIVGGALPAPAVVHGQTAKTITIKVDFKDRKPTSPYIYGTNODLAG 71

Db 4 RQKRLEFVSALAVSLTMTVMPASVNA-AAADVTFTINTQSEAAISPIYGTNODLSG 62

QY 72 DENMAARLGGNRTGYNNENNMSNAGSDWQSQSDNYLCSNGGLTQAECEKPGAVTTSFH 131

Db 63 TENSSRRLLGGNRLTYGYNENNASSAGRWLHYSDDFLCNGGVPTDCKDKPGAVTAFH 122

QY 132 DQSLKLGTYSLVLPNAGYVAKDNGSVQSEKAPSRNQQVYVNAKNAFPQLPDNDNR 191

Db 123 DKSLENGAYSIVTLQWAGYVSRDKNGPVDSETPAPRMDKVEFAKNAPFSLQPHLNDGQ 182

QY 192 VYDFEFVHLVNVKYGTAATKAGVKGYALDNEPALWSTHPRHGEKVGAKELVDRSVLS 251

Db 183 VYDMEEVNVLNRYGNASTGTGKAYSLDNEPALWSTHPRHGEKVGAKELVDRSVLS 242

QY 252 KAYKADAGAEVFGPVLGYFGAYKDLQTAPDMDSVKNGTNSWFVDYVLDQMRSSQVEGKR 311

Db 243 KAVKNDPFAEIEFGPALYFGAYLSLDAPGWSFLOGNSWFDYVLDQMRSSQVEGKR 302

QY 312 LLDVFDVHWYTPAEMGGGIRIT-NEVNDFTKKARMQAPRTLMDPTTYKEDSWIAQWNEEL 370

Db 303 LLDVLDVHWYTPAEMGGGIRIT-NEVNDFTKKARMQAPRTLMDPTTYKEDSWIAQWNEEL 362

QY 371 PILPRLKQSDVKKYYPCTKLAMTEYSYGGENDISGGIAMDVLGILGKNDVYMANVWKLKD 430
 Db 363 PILPRLKQSDVKKYYPCTKLAMTEYSYGGENDISGGIAMDVLGILGKNDVYMANVWKLKD 430
 QY 431 GVNNTVSAAYKLYRNYDCKNSTFGDTSYAQTSIDVSNSSVSHVSTNADSKELHLVWVWKS 490
 Db 423 N-TDYSAAKLYRNYDCKNSTFGDTSYAQTSIDVSNSSVSHVSTNADSKELHLVWVWKS 490
 QY 491 MDSAPDAQFDLSGAKTYISGKVGFDKNSOIKKAPITQISGNRTYTPVPLTAYHVL 550
 Db 482 FDDPINATFQSGDKTYSGRVWGFQDGTGSDTEQAATINNNQFYTTLPLSAYHVL 541
 QY 551 TTGNDTSPV 559
 Db 542 -KADSTEPV 549

RESULT 2
 H97012
 hypothetical protein CAC0915 [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: H97012
 R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97012
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-606 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78891.1; PID:g15023815; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC0915

Query Match 21.68; Score 1552.5; DB 2; Length 606;
 Best Local Similarity 51.3%; Pred. No. 3.1e-81;
 Matches 292; Conservative 92; Mismatches 172; Indels 13; Gaps 7;

QY 12 KRSLWPLVVMACTIVGALPAPAV-VHGQTAKTITIKVDTFKDRKPISYIYGTNDOLA 70
 Db 2 KRNLVLSALATFSSMVLTPVSLGKVPADTA-DVNVNIDTNAEKQALSIYIYGTND 60
 QY 71 GDNWAARRLGGNRTGYNNENNNSAGSDWQSSDNLCSNGGLTQAECEKPGAVTTSF 130
 Db 61 -NAKVTARRIGGNRTGYNNENNNSAGSDWQSSDNLCSNGGLTQAECEKPGAVTTSF 119
 QY 131 HDQSLKLT-YSLVTLPMAGVYKADGNGSVQESKAPSAKRNQVYNAKNAFQLOPLND 189
 Db 120 HDKSLAMGVPSVLTQAGGYVAADQSGPLANTDVAESSKWKVFENKNGPLSLPTD 179
 QY 190 NRVYVDFVHFLVKNKYGTASTKAGVYALDNEPALSHTHPRHGEKVGAKELVDRSVS 249
 Db 180 GSVYDFEVNVLVKNKYGSASSGKIGYSLDNEPSLWPHLHPDKTKCSEVLDKDTQ 239
 QY 250 LSKAVKALDACAEGVPGVLYGFGAYKDLQTA PDWDSVKNYSFVYDYLQMLRSSQVEG 309
 Db 240 LAQVYKKIDPAEATFGPALFGSAFDFNSPDWSSVKNYQWFIYDYLDMKNKNSDAAG 299
 QY 310 KRLLDVDFVHVPYKAMGGGIRI-TNEVGNDETCKKARMAQAPRTLDWPTYKEDSNIAQW 368
 Db 300 KRLLDALDLHWYPAKGGQGVVTSNTSDVNCNKAQAPRLSDMTSTYEDSWIGQCKW 359
 QY 369 FLPLPRLKQSDVKKYYPCTKLAMTEYSYGGENDISGGIAMDVLGILGKNDVYMANVWKL 428
 Db 360 GLPLIPKVKSSIDKYYPCTKLSPSEYNYGGEDHISGGIAQADALYFGKYGVFATYWE 419
 QY 429 KGVNNTVSAAYKLYRNYDCKNSTFGDTSYAQTSIDVSNSSVSHVSTNADSKELHLVW 488
 Db 420 NSDKNNYVQSAFNLYNDGNNKYGDTVKCDTSDINNSSTYASVTSNDGNKMDIIVM 479

QY 489 KSMDSAPDAQFDLSGAKTYISGKVGFDKNSOIKKAPITQISGNRTYTPVPLTAYH 548
 Db 480 KNYTDSINFNVSNNKNTVSGQVWGFDSNNSNITRDDVSSISGNKFTYKIPALTA 539
 QY 549 VLTTGNDTSPVEGPESEFKLKAEGDKVH 577
 Db 540 VLTTAOKSS-VKGDVN-----GDGVN 560

RESULT 3
 A48954
 mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocoellum saccharolyticum
 N:Alternate names: beta-mannanase
 C:Species: Caldocoellum saccharolyticum
 C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
 C:Accession: A48954; B43745
 R: Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
 Appl. Environ. Microbiol. 58, 3864-3867, 1992
 A:Title: The beta-mannanase from "Caldocoellum saccharolyticum" is part of a multidoma
 A:Reference number: A48954; MUID:93119139
 A:Accession: A48954
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-1331 <GIB>
 A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
 A:Note: sequence extracted from NCBI backbone (NCBI:121576, NCBI:121577)
 R: Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
 Appl. Environ. Microbiol. 57, 694-700, 1991
 A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
 A:Reference number: A43745; MUID:91247819
 A:Accession: B43745
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-337, 'PPRQHQRQ' <LUE>
 A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
 A:Note: the authors translated the codon CAC for residue 262 as Glu
 A:Note: this sequence has been revised in reference A48954
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 20.3%; Score 1459.5; DB 2; Length 1331;
 Best Local Similarity 52.2%; Pred. No. 2.4e-75;
 Matches 273; Conservative 83; Mismatches 142; Indels 25; Gaps 5;

QY 32 PAPAVVHGQTAKTITIKVDTFKDRKPISYIYGTNDLAGDENMAARRLGGNRTGYNWE 91
 Db 783 PPSVYVE-----ITINTNAGRTQISPIYIGANQDIEGVVH-SARRLGGNRLTYNWE 833
 QY 92 NNSNAGSDWQSSDNLCSNGGLTQAECEKPGAVTTSFHDOSLKLGTYSVLTPLMAGV 151
 Db 834 NNFSAAGNDWYHSSDDYLCWSMGISGEDAKVPAAVVSKFEHYSKNNAYSATVLMAGV 893
 QY 152 AKDNGSVQESKAPSAKRNQVYNAKNAFQLOPLNDNRVYVDFVHFLVKNKYGTASTK 211
 Db 894 SKDNYGTVENETAPSNRAEYKFKKADAPLSNPLNDNDFVYVDFVHFLVKNKYGTASTK 953
 QY 212 AGVKGVALDNEPALSHTHPRHGEKVGAKELVDRSVLSKAVKALDACAEGVPGVLYG 271
 Db 954 TGKGYLIDNEPALSHTHPRHGEKVGAKELVDRSVLSKAVKALDACAEGVPGVLYG 1013
 QY 272 GAYKDLQTA PDWDSVKNYSFVYDYLQMLRSSQVEGKRLLDVDFVHVPYKAMGGGIRI 331
 Db 1014 MGYSIQDAPDNQVKNYGEHRWFISWLEQMKKASDFKRLLDVDFVHVPYKAMGGGIRI 1073
 QY 332 TNEVGNDETCKK-ARMQAPRTLDWPTYK-----EDSWIAQWNSFELPILPRLKQ 378
 Db 1074 CFDEGNDTSKEVVIARMAQAPRTLDWPTYKSVKGITAGENSINQWESDYLPIPNVKA 1133
 QY 379 SVDKYYPGKTLAMTEYSYGGENDISGGIAMDVLGILGKNDVYMANVWKLKGVNNTVSA 438
 Db 1134 DIEKYPGKTLAMTEYSYGGENDISGGIAMDVLGILGKNDVYMANVWKLKGVNNTVSA 1190

QY 1033 AMSENGIPDPDLMLKAYOAHWSFATWYGDVDRGQNSLEHLKVVYHPNV-ITLDLDP 1091
Db 190 --SKGPIHGICMOCHINWNP-----SVSEIENSILFSSIPGIEIHITELD 234
QY 1092 TNLKTYGITEQPSVPGSFTLNAAGETAKVSLSWTASANAKEYE-VKRSSTTENGAFATVAS 1150
Db 235 MSLYNGSSSENTSTPPQDLLOKQSQRYEFTM-----LKKYKNVVKSVT---FWGLKD 285
QY 1151 D-----VYSSS-----YTDFAVADITYYYQVYVAKNDAGOTLSN-----PASAMPK 1191
Db 286 DYSWLSRFYKGNWPLLFED--YSAKPAYMAVIEA---SGVTSSPTPTPTPTVTVTP- 339
QY 1192 ADTQOQPTTG--LLQYRTADTKVNDHNLNPOQIYVNGKGTSTIPINELKIRYYTIDGPRE 1249
Db 340 TPTPTPGGQIKVLYANKETNSTTIRPWLKVVNSGSSSIDLSRVIWYVTDGERA 399
QY 1250 QTFNCYATLSCSKLNGKLVMEKAATGADYYLEVSNFSDAGVLPAGGSGTDIOIRIHK 1309
Db 400 QSAVSDMAQIGASNTYTFKFKLSSVSGADYYLEIFKSGAGQLQPGKDTGIEIRFNKS 459
QY 1310 DMSNYESDDYSYKGTOTSFADHPKVTLYHNGVLVWGTPTAN 1352
Db 460 DMSNTNOGNDWSLQSWMSYGENEKTAYIDGVLVWGOEPSTN 502
RESULT 8
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Caldocellum saccharolyticum
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
A:Accession: S02711
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A:Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo-
A:Reference number: S02711; MUID:89098398
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1-1039 <SAU>
A:Cross-references: EMBL:X13602; NID:940645; PIDN:CAA31936.1; PID:940646
C:Genetics:
A:Gene: celB
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce-
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1039/Product: cellulase #status predicted <MAT>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 6.1%; Score 441.5; DB 2; Length 1039;
Best Local Similarity 28.6%; Pred. No. 4e-17;
Matches 136; Conservative 71; Mismatches 169; Indels 99; Gaps 19;

QY 943 HEAEGKFWWGAAGPEPVK-----KLYLMHDLRTNVHKLNIWVWNSV-----A 988
Db 129 HMQTPDFFKDSNGORLSKDALLARKQVI-----YDVVGRYKGVYADWVNEAIDENQ 183
QY 989 PD-----WYP--GDEYVD---ILSFDSYPPQA---GDYSPQISKYEDLV-ALGKDKLV 1032
Db 184 PDSYRRSTWYECGPEYIEKAFIWAHEADPAKLFYNDYNTYSEIKKRDFFIYVWVKNLX-- 241
QY 1033 AMSENGIPDPDLMLKAYOAHWSFATWYGDVDRGQNSLEHLKVVYHPNV-ITLDLDP 1090
Db 242 --SKGPIHGICMOCHINWNP-----SVSEIENSILFSSIPGIEIHITELD 286
QY 1091 PTNLKTYGITEQPSVPGSFTLNAAGET-----AKVSLSWTASA 1128
Db 287 DMSLYNGSSSENTSTPPQDLLOKQSQRYEFTMFLKKNVYKSVTFWGLKDDYSWLSRF 346

QY 1129 NAKS-----YEVKRSTTENGAFATV-ASDVYSSYTDTAVTADTYYYOVVAKNDA 1178
Db 347 XGKNDWPLLFEDYSAPKA-----YWAVIASGVTTSSPTPTPTPTV-----VTPTPTP 396
QY 1179 GQTLNNTASAMP-KADTQOQPTTG--LLQYRTADTKVNDHNLNPOQIYVNGKGTSTIPINE 1235
Db 397 TPTPTVATPTPTPTPVSTPATGGQIKVLYANKETNSTTIRPWLKVVNSGSSSIDLSR 456
QY 1236 LKIRYYTIDGDRQTFNCYATLSCSKLNGKLVMEKAATGADYYLEVSNFSDAGVLP 1295
Db 457 VTIRYVTVTDGERAQAQSAVSDMAQIGASNVTFKFKLSSVSGADYYLEIFKSGAGQLQ 516
QY 1296 GGSGTDIOIRIHKTDWNSYNESDDYSYKGTOTSFADHPKVTLYHNGVLVWGTPT 1350
Db 517 GKDTGIEIRFNKSDWNSYNOGNDWSLQSWMSYGENEKTAYIDGVLVWGOEPSTN 571
RESULT 9
T17120
cellulase (EC 3.2.1.4) precursor, thermoactive - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17120; A43745
R:Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A:Title: Cella, another gene coding for a multidomain cellulase from the extreme therm-
A:Reference number: Z18698; MUID:95336703
A:Accession: T17120
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1742 <TEO>
A:Cross-references: EMBL:L32742; NID:9537499; PID:g537500; PIDN:AAA91086.1
R:Luethi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod-
A:Reference number: A43745; MUID:91247819
A:Accession: A43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1516-1544, 'A', 1546-1742 <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72860.1; PID:g144293
C:Genetics:
A:Gene: cels
C:Keywords: glycosidase; hydrolase

Query Match 5.6%; Score 406.5; DB 2; Length 1742;
Best Local Similarity 20.0%; Pred. No. 9.5e-15;
Matches 271; Conservative 171; Mismatches 474; Indels 441; Gaps 57;

QY 89 NWENNMSNAGSDMQSSDNYLCSNGGLTQAECEKPGAVTTSFHDOSLKLCTYSILVTLPM- 147
Db 49 NWVRNWRGDSALKDGDNDGLDTGGWFD-----GDHVKFNLPMS 89
QY 148 -----AGYVAKDG---NGSVOESEKAPSARNQVYNAKNAPFQLQDPLNDNRVYVDE 196
Db 90 YTGTMLSAAAYEKDAFVKSG-----QLEHILNQIEWNDY 125
QY 197 FVHFLVNY-----GTASTKAGVKGYALONEPALWHTHP--RIHGEKVKAGELVDVRSV 248
Db 126 FVKCHPSKYVYVQVGD-----GGKHAWGPAEVMQMERPSFKVTQSSPSASVAETA 180
QY 249 SL-----SKAVKALDAGAEVGPVLYGF-----GAYK 275
Db 181 SLAAASIVLKDNRPTKATYLOHAKO-----LFEFAEVTKSDSGYTAANGYNSWSGYD 235
QY 276 DLQATPDWDSYKGYSWFVDYVYLDQMSQVEGRKLLDVFVHWYPEAMGGGIRITNEY 335
Db 236 ELSWAAVWLYLATNDSTYLTAKESYVQWPKISGSIIDYKWAHCWDVDVHNGAALLAKI 295
QY 336 GNDETTKARMQAPTLNDPTPKED-----SWIAQWNSFELPLRLKQSVKDYKYPG 386
Db 296 TDKDTYKQIIESHLDYWTGTNGYGERIKYTPKGLAWLDQWGLSRVATTATFLAFVYSDWSG 355

QY 1212 VNDHLNPOQIVNKGTTSTIPINELKIRYYTTIDGDEQTFNCNDYATLSCSKLNGKLYKM 1271
Db 699 STTITIRPWLKVNVTSSDLSRVTYTYVDGDKAQAISDWAQIGASNTTFFKVL 758
QY 1272 EKAATGADYILEVFNDSAGVLAPGGSTGDTQIRIHKTDNSNTNESDDYKGTQTSFAD 1331
Db 759 SSSVSGADYILEIGFKSGAQLQAGKDTGEIQIFNKSDNSNTNOGNDWSMWSMNYGE 818
QY 1332 HPKVTLYHNGVLMWGTEPT 1350
Db 819 NVKVTAYIDGVLVWGQEPS 837

RESULT 11
JN0111
cellulase (EC 3.2.1.4) precursor - Bacillus subtilis (strain BSE616)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
C:Accession: JN0111
R:Park, S.H.; Kim, H.K.; Pack, M.Y.
A:Title: Characterization and structure of the cellulase gene of Bacillus subtilis BSE616
A:Reference number: JN0111; MUID:91299280
A:Accession: JN0111
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:D01057; NID:g216387; PID:BAA00859.1; PID:d1001323; PID:g216388
A:Note: The authors translated the codon ATA for residue 102 as Tyr
C:Comment: The low molecular weight of the mature protein suggests carboxyl-terminal processing.
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose
A:Pathway: cellulose degradation
A:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-29/Domain: signal sequence #status predicted <SIG>

Query Match 5.58; Score 396.5; DB 2; Length 499;
Best Local Similarity 24.58; Pred. No. 4.6e-15;
Matches 143; Conservative 81; Mismatches 211; Indels 149; Gaps 23;

QY 783 KTLTNPATYAKALMNYLDQYGNKMLSGQEINE--IDWLQANVGKKPAIALAAL---- 836
Db 47 TQLVNRDCKAVOLKGISGHLQWYG-----EYVNDKSLKWLDRDGGITVFRAMVTAD 99
QY 837 -DLIYSPSRABHLSSTAEKATANDKQGGIVTFAMHWNAPKGLID---TQGREWRGF 892
Db 100 GGIID-NPSVKNMKEAVEAAKEL-----GIYVIDWHI-----LNDGNPNQNKERAKEF 148
QY 893 YADSTTFDEYAMNHPSEDEYKLLIRDIDVIAGQLKLQDAKVPVLRPLRPLHEASGKFWFW 952
Db 149 FKEMSSL---YG-NTPN-----VYIENEPNGDVNWK 177
QY 953 GARGPEPVKLYILMHDRLTNVHKLNNLIWV--NSVAPDYPGDEYVDILSFPSPQAG 1010
Db 178 RDIKPYAEVVISVIRKNDPNDIIVGTCTWSQDVNDAAQDQKQANVMDLHF---YAG 233
QY 1011 DYSPOISKYEDVLVALGKOKKLVAMSENGPIPPDLMKAYQAHWSFWATYWGDFVDRGKQN 1070
Db 234 THGQFL-----RDKANYALSKGAPI-----FVTEWG----- 259
QY 1071 SLEHLKVVNHPNVTLDLPNLKATYGTITEOPSPVGSFTLNAAGETAKV----SLSWTA 1126
Db 260 -----TSDASGNGGVFLDQSRWLYKLDSTKITSVW- 289
QY 1127 SANAKSYEYKRSTTENGAFATVASDVYSSYTDATVADTYYIQ-VVAKNDAGQTLST 1185
Db 290 -----NWNLSQKQESSALKPCASKTGGWRLSD--LSASGTFVRENILGTNDSTKDIPE 342
QY 1186 ASAMPKADTQPTTGLLQYRTADTKVNDHNLNPOFQIVNKGTTTIPINELKIRYYTID 1245
Db 343 ----PAKDPTQENGISVQYRAGDGSMSNQIRPOLKKNNGTNTVDLKDVTARYWYNAK 398

QY 1246 GDEQTFNCNDYATLSCSKLNGKLVKMEKAAATGADYILEVFNDSAGVLAPGGSTGDIQTR 1305
Db 399 -NKQNVDCDYAQLGCGNVTYRFVTLHLPKQGADYILELGFKN--GTLPAGASTGNIQUR 455
QY 1306 IHKTDNSNTNESDDYKGTQTSFADHPKVTLYHNGVLMWGTEP 1349
Db 456 LHNDWSNVAQSDYISFFKSNF-FKTKKTKITLYDQGLKINGTEP 498

RESULT 12
G69593
cellulase (EC 3.2.1.4) bgIC precursor - Bacillus subtilis
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
R:Kunat, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: G69593
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <KUN>
A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13696.1; PID:ell1834
A:Experimental source: strain 168
R:Wackay, R.M.; Lo, A.; Willick, G.; Zuker, M.; Baird, S.; Dove, M.; Moraneli, F.; S
Nucleic Acids Res. 14, 9159-9170, 1986
A:Title: Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.
A:Reference number: A26114; MUID:87066783
A:Accession: A26114
A:Molecule type: DNA
A:Residues: 10-508 <MAC>
A:Cross-references: GB:X04689; NID:g39823; PIDN:CAA28392.1; PID:g39824
A:Experimental source: strain PAP115
A:Note: part of this sequence, including the amino end of the mature form, was confir
R:Lindahl, V.; Aa, K.; Tronsmo, A.
A:Title: Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus subtili
A:Reference number: I40353; MUID:95225656
A:Accession: I40353
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN1>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAAA47429.1; PID:g39777
R:Lindahl, V.; Aa, K.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24239
A:Accession: S24239
A:Molecule type: DNA
A:Residues: 10-291, 'N', 293-508 <LIN2>
A:Cross-references: EMBL:X67044; NID:g39776; PIDN:CAAA47429.1; PID:g39777
A:Experimental source: strain CK-2
R:Wolf, M.; Geczi, A.; Borris, R.
submitted to the EMBL Data Library, December 1993
A:Description: Genes encoding beta glucan-hydrolyzing enzymes in Bacillus subtilis:
A:Reference number: S49103
A:Accession: S49103
A:Molecule type: DNA
A:Residues: 10-508 <WOL>

Query Match 5.4%; Score 388.5; DB 2; Length 499;
Best Local Similarity 24.0%; Pred. No. 1.3e-14;
Matches 143; Conservative 81; Mismatches 222; Indels 149; Gaps 23;

QY 773 PAADRPFAV--TKTLNPNATVEAKALMNYLVDOYGNMLSGOEEINE--IDWLQANVG 828
DB 35 PAAKNGQLSIKGTQLVNRDGVKQVQLGSHGLQWYG-----DFVKNKSLKWLRRDWG 87
QY 829 KKAIAALDLID----YSPRAEHLSTEAEKAIAMDKOGGIYVTFEAWHNAAPKGLID-- 882
DB 88 ITVFRAAMYTADGGYIDNPSPVKNVKEAVEAKEL-----GIVYIIDWHI-----LNDGN 137
QY 883 -TQKEWWRGYADSTFDIEYAMNHPESEYKLLIRIDIDVIAGQLKLODAKVPVLF 941
DB 138 PNOHKEKAKDFKEMSSL---YG-NTPN-----VYIEI 166
QY 942 LHEAEGKWFVGAKGPEVPKLYILMHDRLTNVHKLNNLIWV--NSVAPDPWYPGDEYVD 999
DB 167 ANEPNGDVNWRDIPYAEVVISIRKNDPNIIVGTGWSQDVNDAAADQLKDANVY 226
QY 1000 ILSFDSYPAQAGYSQISKYEDLVAGKDKKLVAAMSENGPIPDPLMKAYQAHWSFATW 1059
DB 227 ALHF---YAGTHGQSL-----RDKANYALS KAPI-----FVTE 257
QY 1060 YGDFVRDQKQNSLEHLKKVYNHPNVIIDELPTNLKTYGTEQSPVPGSFTLNAAGETAK 1119
DB 258 WG-----TSDASGNGGVFELDSQREWLN 279
QY 1120 V-----SLSWTASANAKEYEVRKSTTENGAFATVASDVYSSYTDATVADTYYYQ-VVA 1174
DB 280 YLDSKNISWV-----NNLSDKQESSALPGASKGCGWPLTD--LTASGTFVRENILG 331
QY 1175 KNDAGOTLSNTASAMPKADTQPTTGLLQYRTADTKVNDNLNAPQFOIVNKGTTISPIN 1234
DB 332 NKDSTKERPET---PAQDNPAQAGISVQYKAGDGVNSQIRPOLHIKNNGNATVDLK 387
QY 1235 ELKIRYTYTIDGDRQTCNDYATLSCSKLNGKLVKMEKATGADYILEVSNFSNDAGVLA 1294
DB 388 DVTARYWYNAK-NKGQNFDCDYAQICGGLNTHKFVTLHKPGAGADTYLELGFKT--GTL 444
QY 1295 PGSGTGTDTOTRIHKTQDWSNYESDDYSYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349
DB 445 PGASTGNTQLRLHNDWNSYAGSDYSFFQSNF-PKTKKTLTYHOGKLIWGTEP 498

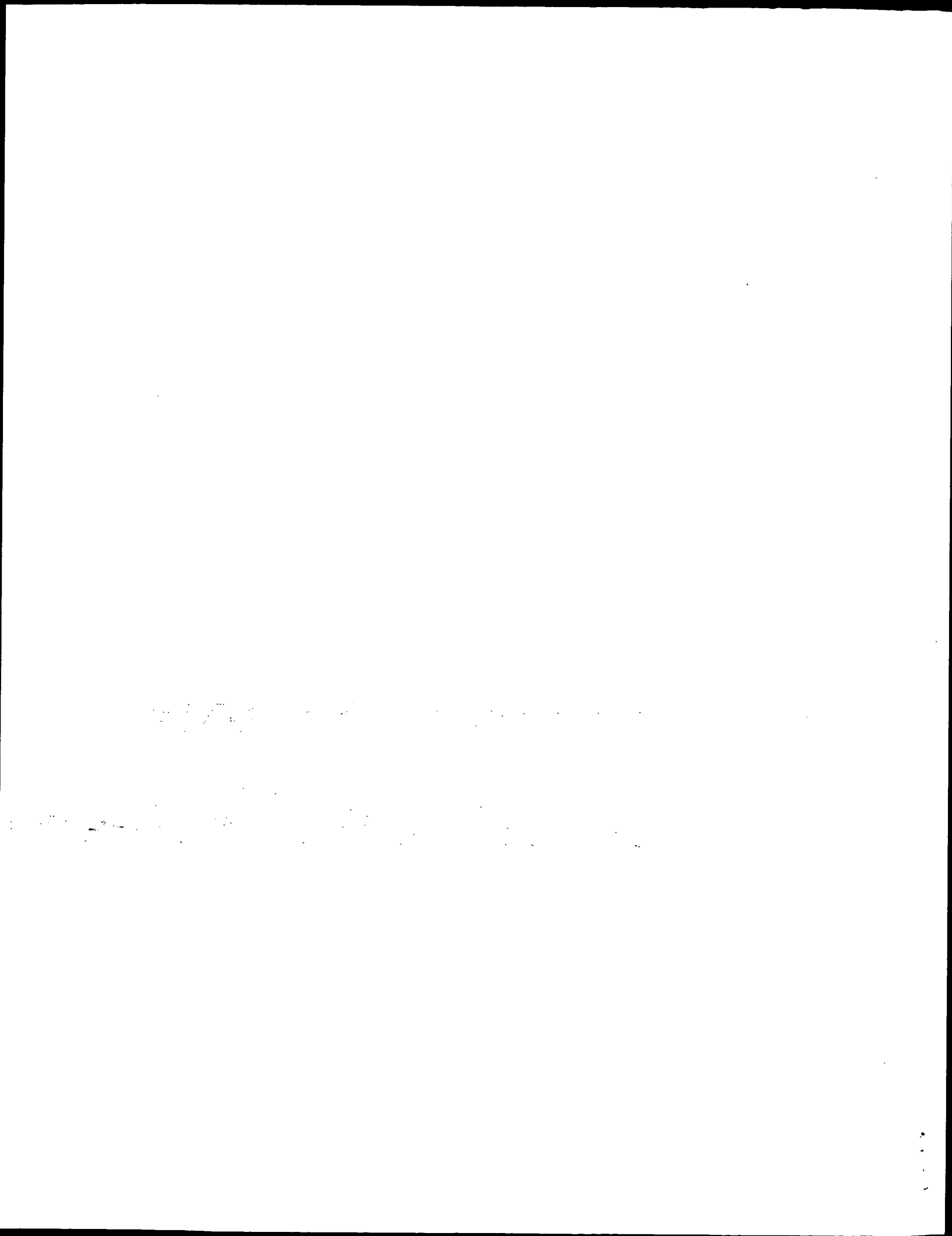
RESULT 15
A41897
cellulase homolog - Bacillus lautus (fragment)
C:Species: Bacillus lautus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-Oct-1999
C:Accession: A41897; S27498
R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.
J. Bacteriol. 174, 3522-3531, 1992
A:Title: ceta from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-1,4
A:Reference number: A41897; MUID:92276330
A:Accession: A41897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <HAN>
A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22302.1; PID:g142662
A:Experimental source: PL236
A:Note: sequence extracted from NCBI backbone (NCBIP:104604)

Query Match 5.2%; Score 371; DB 2; Length 145;
Best Local Similarity 48.3%; Pred. No. 1.8e-14;
Matches 71; Conservative 27; Mismatches 47; Indels 2; Gaps 1;

QY 1203 LQYRTADTKVNDHLMPOQIVNKGTTISIPINELKIRYTYTIDGDRQTCNDYATLSCS 1262
DB 1 LQYRAADTNAADNQIKPSNIRKNGTSYVDLSTLIRYTYTIDGDRQTCNDYATLSCS 60

QY 1263 KLNGKLVKMEKATGADYILEVSNFSNDAGVLA PGSGTGTDTOTRIHKTQDWSNYESDDYSY 1322
DB 61 --NIQISFGNHTGNTSDTYVELSFSEAGSIAAGGSGETQLRMSKTDWMSNEANDYSF 118
QY 1323 KGTQTSFADHPKVTLYHNGVLVWGTEP 1349
DB 119 DGTKTAFADWDVRVLYQNGQIVWGTEP 145

Search completed: July 13, 2002, 00:48:51
Job time: 20147 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:49:05 ; Search time 32.61 Seconds
(without alignments)
1605.300 Million cell updates/sec

Title: US-09-784-554B-2
Perfect score: 7196
Sequence: 1 MRAKNSSNLLFKRSKWLPPV.....PKVTLYHNGVLVWGTEPTAN 1352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1779	24.7	700	1	GUNA_PAEIA	P29719	paenibacill
2	1459.5	20.3	1331	1	MANB_CALSA	P22533	caldocellum
3	937	13.0	1021	1	MANA_RHOMR	P49425	rhodothermu
4	787	10.9	571	1	MANB_PIRSP	P55297	piromyces s
5	768.5	10.7	606	1	MANA_PIRSP	P55296	piromyces s
6	765.5	10.6	569	1	MANC_PIRSP	P55298	piromyces s
7	441.5	6.1	1039	1	GUNB_CALSA	P10474	c endogluca
8	429	6.0	914	1	GUX2_CLOSR	P50900	clostridium
9	406.5	5.6	1742	1	GUNA_CALSA	P22534	caldocellum
10	396.5	5.5	499	1	GUN3_BACSU	P23549	bacillus su
11	394	5.5	499	1	GUN2_BACSU	P10475	bacillus su
12	392.5	5.5	499	1	GUN1_BACSU	P07983	bacillus su
13	374.5	5.2	772	1	CIPB_CLOPM	Q01866	clostridium
14	371	5.2	145	1	YCEA_PAEIA	P29718	paenibacill
15	371	5.2	1853	1	CIPA_CLOPM	Q06851	clostridium
16	366	5.1	986	1	GUN2_CLOSR	P23659	clostridium
17	347	4.8	879	1	GUNI_CLOPM	Q02934	caldocellum
18	329	4.6	505	1	GUNV_ERWCA	Q47096	erwinia car
19	320.5	4.5	504	1	GUNW_ERWCA	Q93935	erwinia car
20	320.5	4.5	1279	1	APU_THESA	P36905	t amylopull
21	316	4.4	1861	1	APU_THETU	P38536	t amylopull
22	315	4.4	1475	1	APU_THETY	P16950	t amylopull
23	291	4.0	513	1	MANB_BACSM	P16699	bacillus sp
24	278.5	3.9	1481	1	APU_THET	P38939	t amylopull
25	258.5	3.6	360	1	MANB_BACSU	P55278	bacillus su
26	258.5	3.6	419	1	MANA_PSEFL	P49424	pseudomonas
27	251	3.5	1848	1	GPBA_CLOCL	P38058	clostridium
28	200.5	2.8	1090	1	GUXB_CLOCL	P50899	cellulomona
29	198	2.8	1902	1	P2P_LACPA	Q02470	lactobacill
30	191.5	2.7	1997	1	PTPB_HUMAN	P23467	homo sapien
31	190.5	2.6	1902	1	P3P_LACLC	P15292	lactococcus
32	189	2.6	1045	1	GUNB_CELFI	P26225	cellulomona
33	188.5	2.6	1433	1	SUBF_BACSU	P16397	bacillus su

34	187.5	2.6	3063	1	CALC_HUMAN	Q99715	homo sapien
35	187	2.6	2334	1	WAPA_BACSU	Q07833	bacillus su
36	186	2.6	1902	1	PIP_LACLC	P16271	lactococcus
37	185.5	2.6	1953	1	BIGA_SALTY	P25927	salmonella
38	185	2.6	972	1	CTA1_BACCI	P94286	bacillus ci
39	182	2.5	1902	1	P2P_LACLC	P15293	lactococcus
40	181.5	2.5	964	1	CTA2_BACCI	P70873	bacillus ci
41	179.5	2.5	865	1	HTRE_ECOLI	P33129	escherichia
42	179	2.5	1597	1	GTF1_STRDO	P11001	streptococ
43	178.5	2.5	1183	1	CNA_STAAU	Q33654	staphylococ
44	178	2.5	1256	1	MRP_STRSU	P32653	streptococ
45	171	2.4	872	1	GUXA_CELFI	P50401	cellulomona

ALIGNMENTS

RESULT	1	GUNA_PAEIA	STANDARD;	PRT;	700 AA.
ID	GUNA_PAEIA				
AC	P29719;				
DT	01-APR-1993 (Rel. 25, Created)				
DT	01-APR-1993 (Rel. 25, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)				
DE	(Cellulase A) (EG-A).				
GN	CELA.				
OS	Paenibacillus lautus (Bacillus lautus).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Paenibacillus.				
OX	NCBI_TaxID=1401;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=PL236;				
RX	MEDLINE=92276330; PubMed=1592807;				
RA	Hansen C.K., Diderichsen B., Joergensen P.L.;				
RT	"Cela from Bacillus lautus PL236 encodes a novel cellulose-binding				
RT	endo-beta-1,4-glucanase."				
RL	J. Bacteriol. 174:3522-3531(1992).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL				
CC	PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-1.				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL				
CC	HYDROLASES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	at the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M76588; AAA22303.1; ..				
DR	PIR; B41897; B41897.				
DR	PIR; S27499; S27499.				
DR	HSP; Q06851; INBC.				
DR	InterPro; IPR001956; CBD_3.				
DR	Pfam; PF00942; CBD_3; 1.				
DR	ProDom; PD001947; CBD_3; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL 1 33 POTENTIAL.				
FT	CHAIN 34 700 ENDOGLUCANASE A.				
FT	DOMAIN 548 700 CELLULOSE-BINDING (BY SIMILARITY).				
FT	ACT_SITE 213 213 BY SIMILARITY.				
SQ	SEQUENCE 700 AA; 3D5C8CADA53EE80F CRC64;				

Query Match 24.7%; Score 1779; DB 1; Length 700;
Best Local Similarity 60.8%; Pred. No. 2.7e-88;
Matches 334; Conservative 75; Mismatches 136; Indels 4; Gaps 4;

QY 12 KRKWLPMVMACTMIVGALPAPAVVHGQTAFTTIKVDTEKRRKPSPIYGTNQDLG 71
 Db 4 RQRKLFYSALAVSLTMTVPMPASVNA-AAADVTFTINTQSERAAISPIYGTNQDLG 62
 QY 72 DENMAARLGNRMGTGYNNENMNSAGSDWQSSDNYLCSNGGLTQAECEKPGAVTTSEH 131
 Db 63 TENSSRLGNRLGTGYNNENMNSAGSDWQSSDNYLCSNGGLTQAECEKPGAVTTSEH 122
 QY 132 DOSKLGTSYLVTLPMAGYVAKDNGSVQSEKAPSRWNOVNAKNAPFQLOPDLNDR 191
 Db 123 DKSLENGAYSIVTLQMGYVSRDKNGPVDESETAPSPRWKVEFAKNAPFQLOPDLNDR 182
 QY 192 YVDFVHFLVKNYGTASTKAGVGYALDNEPALWSTHPRHGEKVGAKELVDVRSVLS 251
 Db 183 YVDFVHFLVKNYGTASTKAGVGYALDNEPALWSTHPRHGEKVGAKELVDVRSVLS 242
 QY 252 KAVKAIDAGAEVFGVLYGFGAYKLOTPADWDSVKNYSWFVYDLDQRLSSQVEGKR 311
 Db 243 KAVKNVDPAELFGAYLGFAYLSLQDAFGWPSLQNGYSWFVYDLDQRLSSQVEGKR 302
 QY 312 LLDVDFVHFLVKNYGTASTKAGVGYALDNEPALWSTHPRHGEKVGAKELVDVRSVLS 370
 Db 303 LLDVDFVHFLVKNYGTASTKAGVGYALDNEPALWSTHPRHGEKVGAKELVDVRSVLS 362
 QY 371 PLPLKQSVKDYKPGCTKLAMTEYSYGGENDISGDIAMTVLGLKNDVYMANVYKLD 430
 Db 363 PLPLKQSVKDYKPGCTKLAMTEYSYGGENDISGDIAMTVLGLKNDVYMANVYKLD 422
 QY 431 GYNNVSAAYLYRNDYDGNKSTFGDTSVSAQTSDIVSNSSVHASVYTNASDKELHLVNMKS 490
 Db 423 N-TDYSAAAYLYRNDYDGNKSTFGDTSVSAQTSDIVSNSSVHASVYTNASDKELHLVNMKS 481
 QY 491 MDSAFDAQFDSLGAKTYISGKVGWFDKNSQIKVDAATSDTENSVAASVYDSESELHLVNLKN 550
 Db 482 FDDPFINATFQSGDKTYISGKVGWFDKNSQIKVDAATSDTENSVAASVYDSESELHLVNLKN 541
 QY 551 TTGNDTSPV 559
 Db 542 -KAUSTEPV 549

RESULT 2
 MANB_CALSA STANDARD; PRT: 1331 AA.

AC P22533;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta-mannanase/endoglucanase A precursor [Includes: Mannan endo-1,4-
 DE beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
 DE mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)].
 GN MANA.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosiruptor.
 OX NCBI_Taxid=44001;
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=9311139; PubMed=1476429;
 RA Gibbs M.D., Saul D.J., Luchi E., Bergquist P.L.;
 RT "The beta-mannanase from 'Caldocellum saccharolyticum' is part of a
 RT multidomain enzyme";
 RL Appl. Environ. Microbiol. 58:3864-3867(1992).
 RN [2]
 RP SEQUENCE OF 1-346 FROM N.A.
 RX MEDLINE=91247819; PubMed=2039230;
 RA Luethi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RT "Cloning, sequence analysis, and expression in Escherichia coli of a
 RT gene coding for a beta-mannanase from the extremely thermophilic
 RT bacterium 'Caldocellum saccharolyticum'";
 RL Appl. Environ. Microbiol. 57:694-700(1991).
 CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH

CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- MISCELLANEOUS: THIS ENZYME IS MOST ACTIVE AT PH 6 AND 80 DEGREES
 CC CELSIUS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L01257; AAA71887.1; -;
 CC EMBL: M36063; AAA72861.1; -;
 CC PIR: B43745; B43745.
 CC PIR: A48954; A48954.
 CC HSP: Q06851; LNEC.
 CC InterPro: IP001956; CBD_3.
 CC InterPro: IP001547; Glyco_hydro_F5.
 CC Pfam: PF00942; CBD_3; 2.
 CC Pfam: PF00150; cellulase; 1.
 CC ProDom: PD001947; CBD_3; 2.
 CC PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Hydrolyase; Glycosidase; Cellulose degradation; Signal;
 CC Multifunctional enzyme.
 KW SIGNAL 1 41
 FT CHAIN 42 1331
 FT DOMAIN 42 325
 FT DOMAIN 326 361
 FT DOMAIN 362 518
 FT DOMAIN 519 564
 FT DOMAIN 565 720
 FT DOMAIN 721 780
 FT DOMAIN 781 1331
 FT ACT_SITE 162 162
 FT ACT_SITE 257 257
 FT CONFLICT 338 338
 FT CONFLICT 340 346
 FT CONFLICT 1331 AA; 146892 MW; FFBCA51BB8D0F0E0 CRC64;
 SQ SEQUENCE 1331 AA; 146892 MW; FFBCA51BB8D0F0E0 CRC64;
 Query Match 20.3%; Score 1459.5; DB 1; Length 1331;
 Best Local Similarity 52.2%; Pred. No. 9.7e-71;
 Matches 273; Conservative 83; Mismatches 142; Indels 25; Gaps 5;
 QY 32 PAPAVVHGQTAFTTIKVDTEKRRKPSPIYGTNQDLGADENNAARRLGNRMGTGYNN 91
 Db 783 PPSVVE-----ITINTAGRTQISPIYGANQDIEGVVH-SARRLGNRLGTGYNN 833
 QY 92 NNNSNAGSDWQSSDNYLCSNGGLTQAECEKPGAVTTSEHDSKLGTSYLVTLPMAGYV 151
 Db 834 NFNENAGNDWYHSSDDYLCWMSGISEDADKVPAAVYKFEYSKKNAYSATVTLQWAGYV 893
 QY 152 AKDNGSVQSEKAPSRWNOVNAKNAPFQLOPDLNDRVYVDFVHFLVKNYGTASTK 211
 Db 894 SKDNYGTSENETAPSNRWAEVKEFKKADAPLSLNPDLNDFVYDFEINLYLNKGNASSP 953
 QY 212 AGVKGVALDNEPALWSTHPRHGEKVGAKELVDVRSVLSKAVKAIDAGAEVFGVLYG 271
 Db 954 TGKLGILNEPDLWSTHPRHGNKVTCKELIEKVEKAVIKTLDPDAEFGVASYGF 1013
 QY 272 GAYKDLQTPDMDSDVSKNGYSWFVYDLDQRLSSQVEGKRLLDVFVHFLVKNYGTASTK 331
 Db 1014 MGYSYLDQADPDWQVKGHRWFISWYLEQMKKASDSFGKRLDLDVLDLWYPEARGNIRV 1073

QY 332 TNEVNDKTK--ARMOAPRTLDPTYK-----EDSWTAQWNSFELPILRLKQ 378
 DB 1074 CFDEGNDTSKEVIAARMOAPRTLDPTYKTSVKQGITAGENSWINQWFSYLPPIPNVKA 1133
 QY 379 SVDKYPTCKLAMEYSYGGNDISGGTAMTDVLGILGKNDVYMANWKLDGYNNTVSA 438
 DB 1134 DIEKYPTCKLAISEFDYGGRNHISGGIALADVLGIFGKGVNFAARW---GDSGSYAAA 1190
 QY 439 AYKLYRNYDGKNSFTGDSYSAQTSADIVNSVSHASVTNASDKELHLVYMKSMDSAFDAQ 498
 DB 1191 AYNILYNDGKSGYGNINVSANTSDVENMPVYASINGQDDSELHIILINRNYDQKLVK 1250
 QY 499 FDLGAKTYISGKVGWFKNSQJKEAPITQISGNRTYTP 541
 DB 1251 INITSPKYRAEYGFDSNPEYKMGNDINIESNVTLEVP 1293
 RESULT 3
 MANA_RHOMR STANDARD; PRT; 1021 AA.
 ID MANA_RHOMR STANDARD; PRT; 1021 AA.
 AC P49425;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78).
 GN MANA.
 OS Rhodothermus marinus (Rhodothermus obamensis).
 OC Bacteria; CFB group; Rhodothermus group; Rhodothermus.
 OC NCBI_TaxID=29549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 43812;
 RA Politz O., Krah M., Borriess R.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -! CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC -! SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X90947; CAA62442.1; -
 DR InterPro: IPR000805; Glyco_hydro_26.
 DR Pfam; PF02156; Glyco_hydro_26; 1.
 DR PRINTS; PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 1021 AA; 115791 MW; A8F7B63109E0F715 CRC64;
 Query Match 13.0%; Score 937; DB 1; Length 1021;
 Best Local Similarity 41.7%; Pred. No. 7.5e-43;
 Matches 204; Conservative 86; Mismatches 155; Indels 44; Gaps 12;
 QY 650 VNGPARYAEETGLAGTIVSSGTCYSGAGYVTFNPNPNDLSLTQIAPTAGLYNLITIG- 708
 DB 476 VAGEIRLEAEDELLGVAVDSTLTGSGRGYVTFGDAEDSDSVRSFEAP-RGVYRVYEGV 534
 QY 709 -----YRSPHDKRTNFSUNGKAFGELLKTKTADPKETSGGKVLNAGANTISFTTG 760
 DB 535 SFSSRPFASYALRVDDWHQTS-----LTKRGGFFEASIGEILWDEGAHTMAFOLM 585
 QY 761 WGYWDYDYLPEAPDRPPHATVTKLTNPNAVTEAKALMNYLVDQYGNMLSGQE----- 815
 DB 586 NG--ALDYVRLPEVSYGPPARPAPQADSQATSAQAALFAFLLSEYGHILAGQQONPYR 643
 QY 816 -EINEIDWLQANVGKKPAALDLIDYSPSRAEHGSLSTEA-EKAIAWDKGGIVTFAMH 873

DB 644 RQFDALNYVRYNTGKPEALVSPDLIDYSPTRAHGVVHYQTPEDWIAGRDGIVSLMWH 703
 QY 874 WNAKPLGIDTQKQ--WWRGFYADSTTFDIEYAMNHPESDYKLLIRIDIVTAGOLKILQ 931
 DB 704 WNAPTDLIEDPSQDCYWWYGFYTRCTTFDVAALADTSSERYRLLLRDIDVIAAOLQKFO 763
 QY 932 DAKVPVPLPLHEAEKWFEMGAKGPEPVKLYILMHDRLTNVHKLNLIWVNSV--AP 989
 DB 764 QADIPVLWRLPLHBAAGFWMGAKGPEPKQLWRLLYERLVHGHGHLNLIWIYTHIEPGA 823
 QY 990 DWYPGDEYDILSFDSPQAGDYSPOISKYEDLVALGDKDKLVAMSENGPIPDPLMKAY 1049
 DB 824 EWYPGAYDIVGRDYYADPD-ALMRSDWNELOTLFGGKLVALTETCTLPDVEVITDY 882
 QY 1050 QAHWSWFAWYGFVDRGKONSLEHLKLYVNHNPVNTLDELPTNLKTYGITQSPSPVGSF 1109
 DB 883 GIWWSFISWTDPFLRDVDP---DRLTRVYHSEVLTTRDELP-DWRSY-----VLHAT 931
 QY 1110 TLNAAGETA 1118
 DB 932 TVOPAGDLA 940
 RESULT 4
 MANB_PIRSP STANDARD; PRT; 571 AA.
 ID MANB_PIRSP STANDARD; PRT; 571 AA.
 AC P55297;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78) (Beta-
 DE mannanase B) (1,4-beta-D-mannan mannanohydrolase B).
 GN MANB.
 OS Piromyces sp.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimastacaceae; Piromyces.
 OC NCBI_TaxID=45796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9631314; PubMed=8768520;
 RA Millward-Sadler S.J., Hall J., Black G.W., Hazlewood G.P.,
 RA Gilbert H.J.;
 RT "Evidence that the Piromyces gene family encoding endo-1,4-mannanases
 RT arose through gene duplication";
 RL FEMS Microbiol. Lett. 141:183-188(1996).
 CC -! CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC -! SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANA, MANC
 CC AND XYNA: AND TO THOSE OF N.PATRICIARUM XYNA.
 CC -! SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X97408; CAA66061.1; -
 DR InterPro: IPR002883; CBD_5.
 DR Pfam; PF02013; CBD_5; 2.
 DR PRINTS; PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase; Signal; Multigene family; Repeat.
 FT SIGNAL 1 19 POTENTIAL
 FT CHAIN 20 571 MANAN ENDO-1,4-BETA-MANNOSIDASE B.
 FT DOMAIN 490 571 2 X 39 AA APPROXIMATE REPEATS.
 FT REPEAT 490 527 1.
 FT REPEAT 533 571 2.

CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; X13602; CAA31936.1; --
 CC PIR; S02711; S02711.
 CC HSSP; Q06851; INBC.
 CC InterPro; IPR001956; CBD_3.
 CC InterPro; IPR001000; Glyco_hydro_10.
 CC InterPro; IPR001547; Glyco_hydro_F5.
 CC Pfam; PF00942; CBD_3; 1.
 CC Pfam; PF00150; cellulase; 1.
 CC Pfam; PF00331; Glyco_hydro_10; 1.
 CC PRINTS; PR00134; GLHYDRLASE10.
 CC PRODOM; PD001947; CBD_3; 1.
 CC PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 CC PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; Repeat;
 CC Multifunctional enzyme; Signal.
 CC SIGNAL 1 28
 CC CHAIN 29 1039 ENDOGLUCANASE/EXOGLUCANASE B.
 CC DOMAIN 376 416 THR/PRO-RICH, TANDEM REPEATS OF T-P.
 CC DOMAIN 417 570 CELLULOSE-BINDING (BY SIMILARITY).
 CC DOMAIN 571 618 THR/PRO-RICH, TANDEM REPEATS OF T-P.
 CC ACT_SITE 177 177 PROTON DONOR (POTENTIAL).
 CC ACT_SITE 285 285 NUCLEOPHILE (BY SIMILARITY).
 CC ACT_SITE 792 792 BY SIMILARITY.
 CC SEQUENCE 1039 AA; 117641 MW; 0B0378171594DDAE CRC64;
 CC
 CC Query Match 6.1%; Score 441.5; DB 1; Length 1039;
 CC Best Local Similarity 28.6%; Pred. No. 3.1e-16;
 CC Matches 136; Conservative 71; Mismatches 169; Indels 99; Gaps 19;
 CC
 CC QY 943 HEAGKWFNGAGKPEPVK-----KLYIMHDLRLTNVHKLNNLWVNSV-----A 988
 CC Db 129 HNOTPDWFFKDSQRLSKDALLARLKQYI-----YDVGRYKGVYANDVYNEADENQ 183
 CC QY 989 PD-----WYP--GDEYD---ILSDSYPOA---GDYSPQISKYEDLV-ALGRDKLV 1032
 CC Db 184 PDSYRSTWYICGPEYIEKAFIWAHEADPNKLFYNDYNTSEIKRDRFIYNNVKNLK-- 241
 CC QY 1033 AMSENGIPDPDLKAYQAHWSFATWYGFVDRDGRQNSLEHLKVVNH-PNV-ITLDEL 1090
 CC Db 242 --SKGIPIGHGQCHINWNP-----SVSEIENSIKLFSSIPGIEHITEL 286
 CC QY 1091 PTNLKTYGTEQSPVSGSFTLNAAGT-----AKVSLSWTASA 1128
 CC Db 287 DMSLYNYGSSNTSTPPQDLLOKQSKYKEIFTMKKYKNVKSVPFWGLKDDYSWLRSF 346
 CC QY 1129 NAKS-----YEKRSITENGATV-ASDVYGSSTDTAVTADTYYQVAKNDA 1178
 CC Db 347 YGRNDWPLLEFDYSAKPA-----YWAVEASGVTTSSPTPTPTPTV-----VPTPTP 396
 CC QY 1179 GQTLNSATAMP-KADTQPTTG--LLQYRTADTKVNDNHLNPOQIVNKGTTISIPINE 1235
 CC Db 397 TPTPTVATPTPTPTVSTPATGQIKVLYANKETNSTTNTIRPLKVVNSGSSSIDLSR 456
 CC QY 1236 LKRYYYTIDGREQFNCDYATLSCSKLNGKLVKMEKAATGADYYLEVFSNDSAGVLAP 1295
 CC Db 457 VTRYTYTIDGREQSAVSDWAQIGASNTYTFKFLSSSVSGADYYLEIGFKSGAGQLQP 516
 CC QY 1296 GGSSTGDIQRIHKTDNSNEDSDYSKCTQTSFADHPKVTLYHNCVLWGTPT 1350
 CC Db 517 GKTGELQIRFNKSDNSNTNQGNDSWLSQMSYGENEKVATYIDGLVWQGPS 571

RESULT 8
 GUX2_CLOSR
 ID AC P50300; STANDARD; PRT; 914 AA.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
 DE (1,4-beta-cellobiohydrolase II) (Avicelase II).
 GN CELV.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1510;
 RN [1]
 RC STRAIN=NCIB 11754;
 RA Bronnenmeier K., Kundt K., Riedel K., Staudenbauer W.H.,
 RA Staudenbauer W.L.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP CHARACTERIZATION.
 RC STRAIN=NCIB 11754;
 RX MEDLINE=91364686; PubMed=1909625;
 RA Bronnenmeier K., Ruecknagel K.P., Staudenbauer W.L.;
 RT "Purification and properties of a novel type of
 RT exo-1,4-beta-glucanase (avicelase II) from the cellulolytic
 RL thermophile Clostridium stercorarium.";
 CC Eur. J. Biochem. 200:379-385(1991).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; Z69359; CAA93280.1; --
 CC HSSP; Q06851; INBC.
 CC InterPro; IPR001956; CBD_3.
 CC InterPro; IPR000556; Glyco_hydro_48.
 CC Pfam; PF00942; CBD_3; 1.
 CC Pfam; PF02011; Glyco_hydro_48; 1.
 CC PRINTS; PR00844; GLHYDRLASE48.
 CC PRODOM; PD001947; CBD_3; 1.
 CC PRODOM; PD011903; Glyco_hydro_48; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
 CC SIGNAL 1 33 POTENTIAL.
 CC CHAIN 34 914 EXOGLUCANASE II.
 CC SEQUENCE 914 AA; 103020 MW; D0DB6017D6DFF82C CRC64;
 CC
 CC Query Match 6.0%; Score 429; DB 1; Length 914;
 CC Best Local Similarity 21.6%; Pred. No. 1.2e-15;
 CC Matches 217; Conservative 110; Mismatches 291; Indels 386; Gaps 42;
 CC
 CC QY 493 SAFDAQFDLSG-----AKTY-----ISGKW-----GFDKNSSQIKAAPIIT 529
 CC Db 150 SNYPSQLDFAVPAGIDPTYNELVSTYGTNTTYGMHLLDNDVNWYGFGRADRISSPAYIN 209
 CC QY 530 --QISGNRTTYVPLTAYHIVLTGTNDTSVPEGPESEFKLAEAGDKVHLSDASSGVV 587
 CC Db 210 TFGSGSBSWETIPQPCW-----DDLITGRNGF-LDLFVGDQO----- 248
 CC QY 588 GYSQVQATDENGPTAVASNLITSTYTDNTVNTGTSYKYVKTAKTNKGSSES--NILKAV 645
 CC Db 249 -YSAQ-----FKYTNPADADARAIAQIATYWANQWAKEHGVNLSQYV 287


```
QY 646 PKMPYNGP-----ARVEAEGTLKGTIVESSGTGYSGAGYVTFNHP-GDSLTMITQIA 697
Db 288 KASRMGDLRYLAMDCKFKIGDSK-----QAGTGYDAHYLSWYANGGGITAD-WA 341
QY 698 PTAGLYNLITIGYRSHDDKRTNFSUNGKAFGELLKTKTADFKETSGGKVLVLNAGANTISF 757
Db 342 WIIGCSHVHAGYONP-----MTAWILANDPEKPSPN-----GAN----- 377
QY 758 EFGWGYDIDYVRLPEADRPAPHATKTLTPNATIVEAKALMNVLYDYGKNNMLSGQBEI 817
Db 378 -----DW-----AKSL-----ERQL 387
QY 818 NEIDWLQANVGKKPAIALDLIDYSPSRAEHLSTSEAKIAWKOQGVTFVFAHWNAP 877
Db 388 EYQWLOS-----REGAIA-----GGAI-----NSY 408
QY 878 KGLDQCKEWRGYFADSTTFDIYAMNHPESEYDKLLIRIDIVIAQLKQLQDAKVPV 937
Db 409 KGRYET-----LPAGISTF-----YGMAYEHPVY----- 433
QY 938 LFRPLHEAGKFWGACGPPEVKLYILMHDR-----LTNVHKLN----- 978
Db 434 -----LDPGSNTWFGQAWTMRQVAEYYLTGTDRAEQLLDKWDWIKSVRLNSDGTPEI 489
QY 979 --NLIVWNVSVAPDWPG-----DEVYDILSF-DSYPOAGDSYPOISKY----- 1019
Db 490 PGNLEW---SGQPDFTWTGTYGNPNLHVSVYSYRTDLGAAGSLANALYYAKTSQDDDEAR 546
QY 1020 -----EDVALGDKDKLVAMSENGPIP----- 1041
Db 547 NLAKELDRMWNLYRDDKGLSAPETREDYVRFEEQEVVYVPOGWSGTMFNGDRIEPVGVTFL 606
QY 1042 -----DPLMKAYQAHWSFWATYWGDFVRDQKNSLEHLKLVYNNPNVITLDELPTN 1093
Db 607 DIRSKYLNDDPYFKLQAY-----NEGR-----APVFNHFWAQCDAIA 647
QY 1094 LKTYGITEQPSVPGSFTLNAAGETAKVSLSWTASANAKEYVKRSTTENGAFATVADSVY 1153
Db 648 NGLYSIL-----FGSEQANDSFITPTSATFDKNQOE-DLSVTVTYNG--NTLLGKIS 696
QY 1154 GSSY-----TDTAVTDTTYYQVAKND--AGQT-----LSNTA 1186
Db 697 GSYLLIEGVYIVNGDV-----IILKEFLAQATGSLILFDPSAGLDRLTITDIITG 751
QY 1187 SAMPKADTQOPTG-LLLYRTADTKVNDNHLNPOFQIVNKGTTSPINELKIRYYTID 1245
Db 752 GGEPEPVEPVEGVVLIQSFNANTQEISNSIMPRFRIYNSGNTSIPLSEVKLYRYTVD 811
QY 1246 GDBEQTFNCYATLSCSKLNGKLVKWEKAATGADYYLVEVSFNSDAGVLAPGSGTGDIQTR 1305
Db 812 GDRQPNFQWCDWASIGSSNVTGTFVKMDGATTGADYILEIGTFPQAGTLEPGASI-EVQGR 870
QY 1306 IHTKMSNTNESDYSYKGTQTSFADHPKVTLYHNGVLVWGTPE 1349
Db 871 FSKIDWTDYTOINDISFNFTASSYVDNFKNITAYISGNLYVIEP 914
RESULT 9
ID GUNA_CALSA STANDARD; PRT; 1742 AA.
AC GUNA_CALSA
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A).
GN CELA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=95336703; PubMed=7612247;
RA Te'o V.S., Saul D.J., Bergquist P.L.;
RT "celA, another gene coding for a multidomain cellulase from the
RL extreme thermophile Caldocellum saccharolyticum";
RN Appl. Microbiol. Biotechnol. 43:291-296(1995).
RP [2]
RX SEQUENCE OF 1516-1742 FROM N.A.
RA MEDLINE=91247819; PubMed=2039230;
RA Leuthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT "Cloning, sequence analysis, and expression in Escherichia coli of a
RL gene coding for a beta-mannanase from the extremely thermophilic
RN bacterium 'Caldocellum saccharolyticum'";
RP Appl. Environ. Microbiol. 57:694-700(1991).
CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
CC CELLULOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
CC SITE FOR PROTEOLYSIS.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L32742; AAA91086.1; -.
CC EMBL; M36063; AAA72860.1; -.
CC EMBL; L01257; -. NOT_ANNOTATED_CDS.
CC PIR; A43745; A43745.
CC HSP; P26221; ITF4.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR000556; Glyco_hydro_48.
CC InterPro; IPR001701; Glyco_hydro_9.
CC Pfam; PF00942; CBD_3; 3.
CC Pfam; PF02011; Glyco_hydro_48; 1.
CC Pfam; PF00759; Glyco_hydro_9; 1.
CC PRINTS; PR00844; GLHYDRLASE48.
CC PRODOM; PD001947; CBD_3; 2.
CC PROSITE; PD011903; Glyco_hydro_48; 1.
CC PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
CC PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
CC SIGNAL 1 23
CC FT CHAIN 24 1742 ENDOGLUCANASE A.
CC FT DOMAIN 24 642 CATALYTIC 1.
CC FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
CC FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
CC FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
CC FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
CC FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
CC FT DOMAIN 1113 1742 CATALYTIC 2.
CC FT ACT_SITE 396 396 BY SIMILARITY.
CC FT ACT_SITE 434 434 BY SIMILARITY.
CC FT ACT_SITE 443 443 BY SIMILARITY.
CC FT CONFLICT 1545 1545 T -> A (IN REF. 2).
CC SEQUENCE 1742 AA; 193696 MW; 3F0699A2123EED07 CRC64;
Query Match 5.68; Score 406.5; DB 1; Length 1742;
Best Local Similarity 20.0%; Pred. No. 4.9e-14;
Matches 271; Conservative 171; Mismatches 474; Indels 441; Gaps 57;
QY 89 NWNNMNSAGSDWQSSDNYLCSNGLTQAECEKPGAVTTSFHDQSLKLTYSILVLP- 147
```


QY 1235 ELKIRYTYIDCDROTFNCNDYATLSCSKLNGKLVKMEKATGADYLYEVSFNSDAGVLA 1294
 Db 388 DTYARYWYKAK-NKGQNFDCDYAQIGCNVTHKEVTLHKPKQGGADTYLELGFKN--GTLA 444
 QY 1295 PGSGTGDIOTRHKTRKNGWYNESDDSYKGTQTSFADHPKVTLYHNGVLYWGTGP 1349
 Db 445 PGASTGNIQLRLHNDWSNYAQSGDYSFFKSNT-FKTTTKITLYDQGLKWGTGP 498

RESULT 12
 GUNL_BACSU STANDARD; PRT; 499 AA.
 AC P07983;
 DT 01-AUG-1998 (Rel. 08, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase).
 GN BGIC OR GLD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DLG;
 RX MEDLINE=87194581; PubMed=3106328;
 RA Robson L.M.; Chambliss G.H.;
 RT "Endo-beta-1,4-glucanase gene of Bacillus subtilis DLG.";
 RL J. Bacteriol. 169:2017-2025(1987).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; M16185; AAA22496.1; ALT_INIT.
 DR PIR; A26874; A26874.
 DR HSSP; O85465; IA3H.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR001547; Glyco_hydro_F5.
 DR Pfam; PF00942; CBD_3; 1.
 DR Pfam; PF00150; cellulase; 1.
 DR ProDom; PD001947; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
 DR KEGG; K01101; Cellulase; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT CHAIN 1 29
 FT SIGNAL 30 499 ENDOGLUCANASE.
 FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
 SQ SEQUENCE 499 AA; 55187 MW; 339D04BE95A63EE1 CRC64;

Query Match 5.5%; Score 392.5; DB 1; Length 499;
 Best Local Similarity 24.5%; Pred. No. 4.6e-14;
 Matches 143; Conservative 78; Mismatches 214; Indels 149; Gaps 23;

QY 783 TKLTNPNATPEAKALNLYVDYGNKWLSSQEEINE--IDWLQANVGKKPAIALDLID 840
 Db 47 TQLVNRGKAVQLGKISHGLQWTF-----DFVKNKSLKWLRDDWGTIVFRAAMYTAD 99
 QY 841 ----YSPSRAEHGLSSTAEKATAMDQGGIVTFAMHNAPKGLID---TQGREWNRGFV 893
 Db 100 GGYIDNPSVKNVKEAVEAAKEL-----GIYVIDWHI-----LNDGNPNQNEKAKEFF 149

QY 894 ADSTTTEDIEYAMNHPESEDEYKLLIRDIDVIAGQLKLODAKVPVFLRPLHEAGKFWFWG 953
 Db 150 KEMSSL---YG-NTPN-----VIVEIANEPNGDVNKKR 178
 QY 954 AKGPEPVKKLYILMHDRLTNVHKLNLIIWV--NSVAPDWPYGEDEVVDILSDFSYPQAGD 1011
 Db 179 DIKPYAEVISVIRKNDPNIIIVGTGTWSQDVNDAAADQLKDANVMYALHF-----YAGT 234
 QY 1012 YSPQISKYEDLVALGDKKLVAMSENGPIPDPLMKAYQAHKSWFATWYGFVRDQKQNS 1071
 Db 235 HGQSL-----RDKANYALSKGAPI-----FVTEWG----- 259
 QY 1072 LEHLKKVYNHPNVIITDELPTNLKTYGITEQPSVPGSFTLNAAGETAKY----SLSWTAS 1127
 Db 260 -----TSDASNGGVFLDQDSREWLNLDSKNISWV-- 289
 QY 1128 ANAKSYEVKRSTENGAFATVASDVYSSYTDYATADTYTYVYQVAKNDAGQTLSNTAS 1187
 Db 290 ----NWLSDKOESSALKPGASKTGWPLTD--LTASGTF---VREING-TKDSTKD 338
 QY 1188 A--MPKADTQPTTGLLLQYRTADTKVNDNHLNPOFQIVNKGTTTIPINELKIRVYYTID 1245
 Db 339 VPETPAQDNPTQEGVSVQYKAGDGRVNSQIRPOLHIKNNGNATVDLKDVTARYWYVK 398
 QY 1246 GDREQTFNCDYATLSCSKLNGKLVKMEKATGADYLYEVSFNSDAGVLPAGSGTGDIOTR 1305
 Db 399 -NKGQNFDCDYAQMGCGNLTHKFVTLHKPKQGGADTYLELGFKN--GTLSPGASTGNIQLR 455
 QY 1306 IHKTKMSYNESDDSYKGTQTSFADHPKVTLYHNGVLYWGTGP 1349
 Db 456 LHNDWSNYAQSGDYSFFOSNT-FKTTTKITLYHOGKLIWGTGP 498

RESULT 13
 CIPB_CLOTM STANDARD; PRT; 772 AA.
 ID CIPB_CLOTM
 AC Q01866;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein SI/SL)
 DE (Cellulose integrating protein B) (Fragment).
 GN CIPB.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS;
 RX MEDLINE=93146373; PubMed=1490597;
 RA Poole D.M.; Morag E.; Lamed R.; Bayer E.A.; Hazlewood G.P.;
 RA Gilbert H.J.;
 RT "Identification of the cellulose-binding domain of the cellulosome
 subunit SI from Clostridium thermocellum YS.";
 RL FEMS Microbiol. Lett. 78:181-186(1992).
 CC -1- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
 CC CELLULOLYTIC ENZYMES PROBABLY THROUGH THE BINDING OF THE NINE
 CC REPEATED DOMAINS WITH THE DUPLICATED SEGMENT OF 24 AMINO ACIDS
 CC PRESENT IN CATALYTIC SUBUNITS OF THE CELLULOSE.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -1- SIMILARITY: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL; X68233; CAA48312.1; -.

[3]
RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
RX MEDLINE=97238934; PubMed=9083107;
RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
RA Frolow F.;
RT "A cohesin domain from Clostridium thermocellum: the crystal
RT structure provides new insights into cellulosome assembly";
RL Structure 5:381-390(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
RX MEDLINE=97076134; PubMed=8918451;
RA Tormo J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
RA Steitz T.A.;
RT "Crystal structure of a bacterial family-III cellulose-binding
RT domain: a general mechanism for attachment to cellulose";
RL EMBO J. 15:5739-5751(1996).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
RX MEDLINE=98022914; PubMed=9402065;
RA Tavares G.A., Beguin P., Alzari P.M.;
RT "The crystal structure of a type I cohesin domain at 1.7-A
RT resolution";
RL J. Mol. Biol. 273:701-713(1997).
CC -|- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
CC CELLULOLYTIC ENZYMES.
CC -|- SUBCELLULAR LOCATION: CELL SURFACE.
CC -|- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
CC COMPONENTS OF THE CELLULOSE.
CC -|- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08665; -; NOT_ANNOTATED_CDS.
DR EMBL; X67506; CAA47840.1; -;
DR PIR; S36859; S36859.
DR PDB; 1ANU; 23-JUL-97.
DR PDB; 1A0H; 08-JUL-98.
DR PDB; 1NBC; 26-SEP-97.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR002102; Cohesin.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00963; Cohesin; 9.
DR Pfam; PF00404; Dockerin_1; 2.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 2.
KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
KW 3D-structure.
FT SIGNAL 1 28
FT CHAIN 29 1853
FT DOMAIN 364 522 CELLULOSOMAL SCAFFOLDING PROTEIN A.
FT DOMAIN 54 1694 CELLULOSE-BINDING (BY SIMILARITY).
FT REPEAT 29 182 9 X 150 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 183 322 1-1.
FT REPEAT 560 704 1-2.
FT REPEAT 724 866 1-3.
FT REPEAT 889 1031 1-4.
FT REPEAT 1054 1196 1-5.
FT REPEAT 1219 1361 1-6.
FT REPEAT 1384 1526 1-7.
FT REPEAT 1548 1690 1-8.
FT REPEAT 1548 1690 1-9.

FT	DOMAIN	1791	1846	2 X 24	AA APPROXIMATE REPEATS.
FT	REPEAT	1791	1814	2-1.	
FT	REPEAT	1824	1847	2-2.	
FT	CONFLICT	1615	1615	A -> AA (IN REF. 1).	
SQ	SEQUENCE	1853	AA; 196831	MW; 3ABDDC03ABFC5372	CRC64;

Query Match 5.2%; Score 371; DB 1; Length 1853;
Best Local Similarity 32.0%; Pred. No. 4.3e-12;
Matches 105; Conservative 48; Mismatches 113; Indels 62; Gaps 12;

QY	1062	DFVRDQKQNSLEHLKVKYNNHPNVTITDELPTNLKTYGITEQP-----SYVPGSF	1109
DB	217	DFVRYDPNVLE---IIGIDPGDIIVDPNPT--KSFDTAIYDPKRIIVFLFAEDSGTGAY	271
QY	1110	TLNAAGETAKVSLSWTASAKSY-----EVKRSTTENGAFATVADSVYG	1154
DB	272	ATYKDGVFAKIRA--TVKSSAPGYIIFDEVGGFADNDLVEQKVSFIDGGVNVGNATPTKG	329
QY	1155	SSYTDTAV---TADTTYVYQVAKNDAGQTLNSTASAMPKADTQQPTTG-LLLQYRTADT	1210
DB	330	ATPTNTATPTKSATATPTSPSVPTNPTNPANT-----PVSGLAKVEFYNSNP	378
QY	1211	KVNDNHLNPOFOIVNKGTTSPINELKIRYYTIDGDRBOTFNCDYATL-----SC	1261
DB	379	SDTNSINPOKVTNTGSSAIDLSKLTLYRYTVDGQKQDTFCDHAAIITSGNSYNGIT	438
QY	1262	SKLNGKLVKMEKAATGADYYLVEVSENSDAGVLAPGSGTGDIQTRIHKTDWSYNESDDYS	1321
DB	439	SNVKGTFVKMSSNTNADTYLEISFT--GGTLEPGAHV-OIOGRFAKNQMSNTYQSDNYS	495
QY	1322	YKGTQTSFADHPKVTLYLHNGVLVWGTEP	1349
DB	496	FK-SAQFVEWDQVTAIYNGVLVWGKEP	522

Search completed: July 13, 2002, 00:59:23
Job time: 618 sec

Query Match 23.5%; Score 1688; DB 2; Length 1601;
Best Local Similarity 43.9%; Pred. No. 2.8e-76;
Matches 359; Conservative 93; Mismatches 246; Indels 120; Gaps 16;

QY 29 GALPAPAVVHGOT--AKTITIKVDTFKORCKPISPIYIYGTNODLAGDENNAARRLGGNRT 86
DB 760 GRAPEGVTPNTEPAKVVDIIRIDITSABERKPISPIYIYGSNQL--DATVAKRFGGNTT 817
QY 87 GYNWNNMNSAGSDNQSSDNYLCSNGGLTOAECSEKPGAVTTSFHDQSL-KLGYSLVTL 145
DB 818 GYNWNNMNSAGSDMLHYSDTYLLEDGGVPGKGSWTPASVVTTFHDKALSNVPTLTL 877
QY 146 PMAGYVADKNGSVQSEKAPKARNQVNNAKNAPQLOPLNDNRVYVDFVHFLVNY 205
DB 878 QAAGYVSADGNGPVSQSEETAPSRKVEKFEKGPFSLTPTDDEYVYMDFVNYLVN 937
QY 206 GTASTKAGVKGVALDNEPALWSHTPRTHGEKVGAKELVDRSVLSKAVKAIDAGAEVFG 265
DB 938 GNASTPTGKIGYSDNEPALWSHTPRHPDNVTAKELIEKSVALSKAIVKVDYAEIFG 997
QY 266 PVLXGFGAYKDLQATPDWDSVKGNYSWFVDYLDQMRLSSQVEGRLLDVPDWHVYPEAM 325
DB 998 PALYGFAYETLQSPADWNGTEGEGYRWFIDYVLDKMKKASDEEGKRLLDVLDVHYPEAR 1057
QY 326 GGGIRIT--NEVGNDETKKARMAQPTLWDPTYKEDSHIAQWNSEFPIPLRLKQSDVKY 383
DB 1058 GGERICFGADPRNIETNKAQLQAPRTLMDPTYIEDSWIGQWKDFPLPILNLLDSIEKY 1117
QY 384 YPGTKLAMTEXSYGGENDISGGIAMTDVLGILGNNDVYMANVYKLDGCVNYSAAVKLY 443
DB 1118 YPGTKLAITEYDGGNHTGGIAQADVILGIFGKYGYLATFW--GASNNYTAGINLY 1175
QY 444 RNYDKNSTFGDTSVSAQTSQDIVSNVSHASVTNADSKELHLDVNMKNSMDSAFDAQFDLSG 503
DB 1176 TNYDGKGGKFGDTSYKCTSDIEVSSAYASIVGEDSKLHILLNKNYDQPTTFNFSDS 1235
QY 504 AKTISGKVGWGFDRKNSOIKEAPITQISGNRETYTYPLTAYHIVLTG-----ND 555
DB 1236 SKNTYIGNWAFDRGSSNITORTPLVINKDNTFTYTPALTACHIVLEAAEPVYVYGLNN 1295
QY 556 TSPVEGPESFKL-----AEAGDKVHLSWDASSGVVGVYVOR-----593
DB 1296 DSKVNAVDMKRYILGIDINLTAAD-----IYFGVNVSSDYNIMKRYLLKAIEDIP 1351
QY 594 -----ATDNGPFTAVASNLTE-----TSYDTNTVNGT 622
DB 1352 YVPENQAPKALFTSPDPVTDENVVNA-SNSIDEDGTIAYAWDFGDCYEGTSTPTTI 1410
QY 623 SYYY-----KVTAKTNKGSSESNILKAVPKMPVNGPARYEAEGETLKGITIVESG 674
DB 1411 TYKYKNPGTYKVLIVTDNQASSFTATIKVTSATGDNKFNEDGTLCG--FTTSGT- 1467
QY 675 YSGAGYVNFN-----PGSLTMTIQAAPTAGL 702
DB 1468 -NATGIVRSPHDDRTNFSLNGRAFGLLKKTKADFKE 740
QY 703 YNLITIGRSPHDDRTNFSLNGRAFGLLKKTKADFKE 740
DB 1527 IAAIQYIMPTPDWSEVLNWNSTWKGTYMV-KTDDWNE 1563

RESULT 2
Q977Y3 ID Q977Y3 PRELIMINARY; PRT; 606 AA.
AC Q977Y3
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ENDOGUCANASE A (ENDO-1.4-BETA-GLUCANASE) (CELLULOSE A),
DE SECRETED, DOCKERIN DOMAIN.
GN CAC0915
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

Clostridium.
NCBI_TaxID=1488;
[1]
SEQUENCE FROM N.A.
RC STRAIN=AFCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: A607607; AAK78891.1; -;
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR02048; EF-hand.
DR Pfam: PF00404; Dockerin_1; 2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 606 AA; 66778 MW; EC00E082846F4CDB CRC64;

Query Match 21.6%; Score 1552.5; DB 16; Length 606;
Best Local Similarity 51.3%; Pred. No. 4.4e-70;
Matches 292; Conservative 92; Mismatches 172; Indels 13; Gaps 7;

QY 12 KRSLKPLVPMACMTMIVGGALPAPAV-VHGOTAKTITIKVDTFKORCKPISPIYIYGTNODLA 70
DB 2 KRKILVSLAIAFAFSSWMLTPVSGLVAFADTA-DVNVNIDTNAEKQAIPIYIYGTNODFS 60
QY 71 GDNMAARRLGGNRMTGYNNMNNMNSAGSDNQSSDNYLCSNGGLTOAECSEKPGAVTTSF 130
DB 61 -NAKVTTARRIGGRSTGYNNMNNMNSAGSDNQSSDNYLCSNGGLTOAECSEKPGAVTTSF 119
QY 131 HDOSLKLGT-YSLVTLPMAGYVADKNGSVQSEKAPKARNQVNNAKNAPQLOPLND 189
DB 120 HDKSLAMGYPYSLVTLQAGGYVAADQSGPLANTDVPASKMKKVEFNKNGPLSLPTDPTD 179
QY 190 NRYVDFEYHFLVNYKGTASTKAGVKGVALDNEPALWSHTPRTHGEKVGAKELVDRSVS 249
DB 180 GSYVDFEYHFLVNYKGTASTKAGVKGVALDNEPALWSHTPRTHGEKVGAKELVDRSVS 239
QY 250 LSKVAIDAGAEVFGVLYGFCAYKDLQATPDWDSVKGNYSWFVDYLDQMRLSSQVEG 309
DB 240 LAQVWKIDPAAETFGPALFSGFAFNDSPPDWSSVKGNYQWFIIDYLDNMKNKSDAAG 299
QY 310 KRLLDVDFVHVPKAGGGGIRT-TNEVGNDETKKARMAQPTLWDPTYKEDSHIAQWNSE 368
DB 300 KRLLDALDHLNTPKAGGGGRTTSDTSNVDCNKARMAQPTLWDPTYKEDSHIAQWNSE 359
QY 369 FLPLRLKQSDVKYYPGTKLAMTEYSYGGENDISGGIAMTDVLGILGNNDVYMANVYK 428
DB 360 GLPLPKVKSSIDKYYPGTKLFSFSEYNYGGEDHISGIIAQADALGVFCYGVYFATYWE 419
QY 429 KDCVNNYSAAKLYRNDYDKNSTFGDTSVSAQTSQDIVSNVSHASVTNADSKELH 488
DB 420 NSDKNNYVQSAFNLYNNYDGNNSKYGTDVCKDTSIDNNSSTYASVTSNDGKNMDI 479
QY 489 KSMDSAFDAQFDLSGAKTVISGKVGFDKNSOIKEAPITQISGNRETYTYPLTAYH 548
DB 480 KNYTDSINFNFNVSNNKNTYSCQVWGFDSNNTKRDVSSISGNKFTYKIPALTAVH 539
QY 549 VLTGNDTSPVEGPESFKLKAEGADGKVH 577
DB 540 VLTTAQKSS-VKGDVN-----GDGVN 560

RESULT 3
Q9A0G4 ID Q9A0G4 PRELIMINARY; PRT; 1751 AA.
AC Q9A0G4
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 CELE.
 DE Caldicellulosiruptor sp. Tok7B.1.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoaerobacter group; Caldicellulosiruptor.
 OX NCBI_TaxID=80339;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TOK7B.1;
 RX MEDLINE=20171169; PubMed=10706665;
 RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
 Bergquist P.L.;
 RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
 thermophile Caldicellulosiruptor isolate Tok7B.1.";
 RL Curr. Microbiol. 40:333-340(2000).
 DR EMBL: AF078042; AAK06394.1; -;
 DR HSSP: Q06851; INBC.
 DR InterPro: IPR001956; CBD_3.
 DR InterPro: IPR001701; Glyco_hydro_9.
 DR InterPro: IPR000566; Lipocin_cytFABP.
 DR InterPro: IPR002965; P-rich_extensn.
 DR Pfam: PF00942; CBD_3; 1.
 DR Pfam: PF00759; Glyco_hydro_9; 1.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR ProDom: PD001947; CBD_3; 3.
 DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 21.2%; Score 1524; DB 2; Length 1751;
 Best Local Similarity 52.9%; Pred. No. 5.4e-68;
 Matches 283; Conservative 83; Mismatches 145; Indels 24; Gaps 6;

QY 32 PAPAVVHGTAQITIKVDFKDRPISPPIYGTQDLAGDENMAARRLGGNRMGTGYNWE 91
 DB 1225 PTPAV-----TPDVKISIDTSRGRTKISPIYIGANQDIGVVH-PARRLGGNRLGYNWE 1278
 QY 92 NNMNAGSDQWQSSDNYLCSNGLTQAECEKPGAVTTSFHDOSLKLGYSLVTLPMAGYV 151
 DB 1279 NNMNAGSDWYHSDDMYIMGITGNDKNVPAAYVSKFHEQSIQONAYSATLQWGVV 1338
 QY 152 AKDNGSVQESKAPSRQNVVNAKNAPFOLQPLDNRVVDVEFHLVKNKYCTASTK 211
 DB 1339 AKDNGTVSESETPSPRAEVEKFKDGALSQPDVNDNYVDMDEFINLKNYGRSSA 1398
 QY 212 AGVKGVALDNEPALMSHTPHRIHGEKVGAKELVDRSVLSKAVKAIDAGAEEVGPVLXGF 271
 DB 1399 TGIGKYLNEPDLWFTTHPHIPKQVTCSELINKSVELAKVIKTLDPDAEIFGPASGYF.1458
 QY 272 GAYKDLQAPDWSVKNYSWFVYLDQMRLLSQVEGRLLDVPDVHVPAMGGGTRI 331
 DB 1459 VGLTLQDAPDQNVQNHFWLSYLEQMKKASDSFGKRLDVLVDIHWPEAQVGGVRI 1518
 QY 332 TNEVGNDERTKK-----ARMAQPTLMDPTYK-----EDSWIAQNNSEPLTLPLK 377
 DB 1519 CFD-GENSTSRVATARMQAPTLMDPTYKTKQGITAGENSWINQWPEYLPPLPNIK 1577
 QY 378 QSVDKYPTGKLTAMTEYSYGGENDISGGTAMTDVILGKNDYVMANYWKLKDGYNVTVS 437
 DB 1578 ADIDKYPTGKLTAEFDYGGKHISGGIALADVLGFGKYGVMYAAW---GDSSGYAQ 1634
 QY 438 AAKLYRNDGKNSFGDTSVAQSDIVNSSVHVASVNASKEHLVVMNKSMDASDA 497
 DB 1635 AAYNIYLDGKSGRSGTSCVSAETDVENMPVYASIEGEDSDTTHIILINNYDKRLKA 1694
 QY 498 QFDLSGAKTYISGVKMGFDKNSQIKAAPIQISGNRTYTVPPTLTAHVHLTT 552
 DB 1695 EIKMNNTRYTGGEIYGFDTSSQIRKMGVLSNIONNTITIEVPLTYHVHLTS 1749

RESULT 4

Q30654
 ID O30654 PRELIMINARY; PRT; 398 AA.
 AC O30654;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE BETA-MANNANASE (EC 3.2.1.78).
 GN MANA.
 OS Dictyoglomus thermophilum.
 OC Bacteria; Dictyoglomus group; Dictyoglomus.
 OX NCBI_TaxID=14;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R446B.1;
 RX Reeves R.A., Gibbs M.D., Bergquist P.L.;
 RT "Sequence of a beta-mannanase from the extreme thermophile
 Dictyoglomus thermophilum R446B.1.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF013989; AAB82454.1; -;
 DR InterPro: IPR000805; Glyco_hydro_26.
 DR Pfam: PF02156; Glyco_hydro_26; 1.
 DR PRINTS: PR00739; GLHYDRLASE26.
 DR Hydrolase; Glycosidase.
 KW SEQUENCE 398 AA; 46492 MW; 0E564DEBA52CA18B CRC64;
 SQ
 Query Match 16.0%; Score 1148.5; DB 2; Length 398;
 Best Local Similarity 54.8%; Pred. No. 4.4e-50;
 Matches 218; Conservative 53; Mismatches 120; Indels 7; Gaps 5;
 QY 702 LYNLTIGVSPHDDKRTNFSLNGKAFGELLKKTADKETSGGKVLNAGANTISFETGW 761
 DB 1 MHELIIGAAAYGYKENSLYNGEFQTNVFPQSQKFTTVYAGLIPLKNGKNTISIVKSW 60
 QY 762 GWYDIDYVRLPEAPDRPHAVTKLTNPNAIVEAKALMNYLVDOYGNMLSGQEEINEID 821
 DB 61 GWELLDYFKIK-KAEIPTMNPNTKLVTPNSKEAQKLMYLSYGYTSLSGOMGYKDAF 119
 QY 822 WLQANVGKKPAIALDLIDYSPRAEHLGSLTEAKAIANDKGGIVTFFAHWNAPKGLI 881
 DB 120 WIMNITDKFPAICGDMYDYSRVERGASSRDVEDAIDWMNMGIVQFQWHNAPKGLY 179
 QY 882 DTQCKEWRGYPVADSTTFDIEYANNHPESDYKLLIRDIDVIGALKKLODAKVPVLP 941
 DB 180 DTPCKEWRGYPVATSTFIEYALNHPESDYKLLIRDIDVIGALKKLODAKVPVLP 239
 QY 942 LHEAEGKFWMGAKGPEPVKKLYILMHDRLTNVHKLNNLWVNSV-APD---WYPGDEY 997
 DB 240 LHEAEGRWFWMGAKGPEACKLWELLDRLVNYHKLNNLWVNTTDSPDALKWYPGDEY 299
 QY 998 VDILSFDISYPOAGDYSPQISKYEDLVALGDKKLVAMSENGPIPPDLMKAYQAHWSFA 1057
 DB 300 VDIYGADYILKDKDYSPSTGMFYNIIVLFGGKLVALTENGIIPPDLMKEQKAYWVWF 359
 QY 1058 TWYGDVFDGDKQNSLEHLKVVNHPNVTITDELPTNLK 1095
 DB 360 TWSG-FENDPNKNEISHIKKVFVHPFVITKDELP-NLK 395
 RESULT 5
 Q934F9
 ID Q934F9 PRELIMINARY; PRT; 808 AA.
 AC Q934F9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE CELLULASE PRECURSOR.
 GN ENDB.
 OS Ruminococcus flavefaciens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Ruminococcus.
 OX NCBI_TaxID=1265;
 RN [1]

Db 408 GGEPIISGILNLTNYDGGKCGCFDGLIPASTEDVSKSSYAAVNDGDESKVTVMITKN 467
Qy 491 MDSAFDAQFDLSGA-KTYISGKVMGFDKNSSQIKAEAPITQISGNRTYVPLTAYHIV 549
Db 468 MTEAENAVIDLKASDKYSAAYVAVYDNDQVRLDVKDKNVNVELPAFSAAMVV 527
Qy 550 LTTGNDTSPVEGSPFKLAEAGDGKVLHSLWDASSGVVGVSVQRAIDENGPFPTAVASNL 609
Db 528 VS--DAAAFDGEKI--YEEKVKTEKTEEFKDPSSNIN 561
Qy 610 ETSYTDNTVNTGTSYVYKVTAKTNKSSSILKAVKMPVNGPARVEAEEGLKGTIVE 669
Db 562 KNGVEIPITDP-----EQVSKIIVN-----GDVTS 587
Qy 670 SSGTGYSGA 678
Db 588 SAGSGWATA 596
RESULT 7
P77847
ID AC P77847 PRELIMINARY; PRT; 911 AA.
DT 01-FEB-1997 (TREMREL. 02, Created)
DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE BETA-1.4-MANNANASE (EC 3.2.1.78).
GN MANA.
OS Caldicellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RT8B.4;
RX MEDLINE=96291679; PubMed=8764509;
RA Gibbs M.D., Elinder A.U., Reeves R.A., Bergquist P.L.;
RT "Sequencing, cloning and expression of a beta-1,4-mannanase gene,
RT manA, from the extremely thermophilic anaerobic bacterium,
RT Caldicellulosiruptor R8B.4";
RL FEMS Microbiol. Lett. 141:37-43(1996).
DR EMBL: U39812; AAC44232.1;
DR InterPro: IPR000805; Glyco_hydro_26.
DR Pfam: PF02156; Glyco_hydro_26; 1.
DR PRINTS: PR00739; GLYDRLASE26.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 911 AA; 103324 MW; AD4A7FF0C1C8DF0B CRC64;

Query Match 14.0%; Score 1010; DB 2; Length 911;
Best Local Similarity 33.6%; Pred. No. 1.3e-42;
Matches 262; Conservative 88; Mismatches 250; Indels 180; Gaps 24;
Qy 379 SYDKYYPGKLTAMTSYSGGENDISGGIATDVLGILKNDVY-----MANYMKLDGV 432
Db 237 AVEDFTGKLVYPEQ-----GKSFYNFEIDTMGFYSGDGF 275
Qy 433 NNVV-SAAVKLRVNDGNSTFGDTSVSAQSDIVNSSVHVASVTNASKELHLVNMKSM 491
Db 276 NKKTSLEYSDLKRSNGNAL-----KVNASLAGTAFDEMNIATKLTDK 320
Qy 492 DSAFAQFDLSGAKTY-----ISGVY-----WGFKNSSQI--KEAAPIT 529
Db 321 D---DKKDFSKYSTLEYTLYPNPKISGKLVASADNPWQIITKDTALNYKDKNAIQ 377
Qy 530 QISGNRTYVPLTAYHI-----VLTGNDTSPVEGP-----ESFKLAEAGDGKVLHLS 579
Db 378 KINGDYAVIKSDNLYNNTKANVLRLRIAGSYVYKTYGPIYDINKLVA---GK---- 429
Qy 580 WDASSGVVGVSVORADENGPTAVASNLTSYTDNTVNTGTSY--YKVTAKTNKGSSE 638
Db 430 -----KVAPKVKTS-----SIPNKNYRVKIEAET----- 456

Qy 639 SNILKAVPKMPVNGPARVEAEEGLKGTIVESSGTGSGAGVYVTFNHPNPGDLSLTMTIOAP 698
Db 457 -----AKGNWYS--IERENAKISGKGYLLIFGNMGTLYDLKIP 495
Qy 699 TAGLYNLTI-----GYRSPHDKRTNFSLNGKA-----FCELLKKTADFKETS 742
Db 496 KTGYYITFISSTLTGMVNYGSV-DIWDIGELKGAAPVNVKGFQEVVVMK----- 545
Qy 743 GKVLLNAGANTISFTGTGNGW-YDIDYVRLP--AADRPPHAYVTYTLNPNPNAIVEAKALM 799
Db 546 --KIYLTQGEHTLSLQKSGGYTIAIDYFTIEELVLANKNKISVDTKLVTNPNPNAQKLM 603
Qy 800 NYLVDQYGNMLSGQ--EINEIDWLQANVGKPAALADLIDYSPSRAEHLSTAE 856
Db 604 KYLASIYGEKILSGQSSGDKREIQMIFDVTTRYPAVRGDFDMDSYSPSRVEHGTGTDVE 663
Qy 857 KATANDKGGIYTFAWHNAKGLIDTQGEKWRGFGYADSTTFDIEYAMNHPESEDKLL 916
Db 664 EAIKWKSGGIVAFCHWNAPTGLIDQCKEWRGFGYTEATTFDIKAMDNPNSTEYKLI 723
Qy 917 IRDIDVIAQKKLQDAKVPVLPRLPHEAGKFWGAKGPEPVKLYILMHDRLTNVHK 976
Db 724 LRDIATAEQLRQKEGVPVLPRLPHEASGGFWGAKGPEPYIKLWKLDFDLVNYHK 783
Qy 977 LNNLIWVNSVAPDWYDGYDILSFDYSYPOAGDYSPOISKYEDVLVALGKOKKLVAMSE 1036
Db 784 LNNLIWNGQDAAYPGDNYVDIIGEDIYEKAQISPYADRFPAKALKYTNARKMIALTE 843
Qy 1037 NGPIPPDMLKAYQAHWSWFATWYGFVRDQK-----NSLEHLKKVYNHPNVTILDEL 1091
Db 844 CGTIPDPAVLKQSGISWFSWAGNFWMTGSKYNDENHMLKIYNSDYVITKDLP 903
RESULT 8
Q9F1T9
ID AC Q9F1T9 PRELIMINARY; PRT; 591 AA.
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE MANNANASE 26B (EC 3.2.1.78).
GN MAN26B.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC Kurokawa J., Hemjinda E., Arai T., Karita S., Kimura T., Sakka K.,
RA Ohmiya K.;
RT "Sequence of the Clostridium thermocellum Mannanase Gene man26B and
RT Characterization of the Translated Product";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB044406; BAB19050.1;
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000805; Glyco_hydro_26.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF02156; Glyco_hydro_26; 1.
DR PRINTS: PR00739; GLYDRLASE26.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 591 AA; 67048 MW; 92AFCF808BD761EB CRC64;

Query Match 12.1%; Score 874; DB 2; Length 591;
Best Local Similarity 38.5%; Pred. No. 4.5e-36;
Matches 214; Conservative 88; Mismatches 188; Indels 66; Gaps 19;
Qy 655 RYAEAGTLKGTIVESSGTGSGAGYTN-----FHNPGDLSLTMTIOAPTAGLYNLT 706
Db 37 KYEEDGIIHGAQIYTDYVGONEYGEVFDLTGTSCSPIAOKGTSTSVNVEDKGLYEIF 96

```

QY 707 IGYRSPHD-DKRTNF-SLNGKAFGELLKKKTADPKETSGGKVLNAGANTISFETGNGWY 764
Db 97 ICYVQYDKKKVQYLVNNGVNGQGEISFPFTLKWREISAGIVKLNAGINIELESYGYT 156
QY 765 DIDYVRLEPAADRPDH-AVTKTITNPATVEAKALMNYLVDOYGNKMLSGQEEI----- 817
Db 157 YFDYLVKPADESIVELKVPKLVNPNATVEAKALMSYLVDIYGHILSGQEEICGSHNY 216
QY 818 ----NEIDWLQANVGKKPAIALDLIDYSPSRAEHGLSSTE--AEKAIAMDQ--GGIVTF 870
Db 217 PGSEAEFTYIOEKTGKLPVAVRGDFMNY-----RGNGLMWDQCAERVIEWYKKGGIPTV 272
QY 871 AWHNAPKGLIDTQGEKWRGFEYADSTFTDIEYAMNHPESEYKLLIRIDIVIAQLKKL 930
Db 273 CHWFSFGDI-----GKADNSFTTSTFISRALT-PTGEENIALALLDIDTIAKRLQV 327
QY 931 QDAKVPVLPRLPHEAGKFWGAGKPEPKKLYILMHDRLTNVHKLNNLIWVNS----- 986
Db 328 QDAKVPVLPRLPHEAGKFWGAGKPEPKKLYILMHDRLTNVHKLNNLIWVNS----- 986
QY 987 VAPDWYPGDEYVDILSFSYPOAGDYSPQ-----ISKYEDLVAGKDKKLVAMSENGPIPD 1042
Db 388 TSAAMYPGDGVVDIIGYDKY-NAKDGKPNGSAISSTFYNLVKTNGKLVAMTENDTIPR 446
QY 1043 PDLMKAYQAHWSWATWYGFVDRGCKNSLEHLKKVYNHNPVITLDELPTNLKTYGITE- 1101
Db 447 VSNLVNERAGWLYFCPWGWLITSEQNPNPVDWLVEMVQSDYCIITLDELPTNLKTYGITE- 1101
QY 1102 -----QPSVPGSET-----LNAAGETAKVSLWSANAKSEYKRVSTTENGAFATVA 1149
Db 506 EDSNPDPSPTPTQPKITYGDLNGDKVNSTDLTIMKRYILKNFD-KLAVPEE-----A 558
QY 1150 SDVYGS---SYTDNAV 1162
Db 559 ADLNGDGRINSTDLSI 574

RESULT 9
ID Q9REC7 PRELIMINARY; PRT: 589 AA.
AC Q9REC7;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE MANNANASE A PRECURSOR (EC 3.2.1.78).
GN MANA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN-VS;
RX MEDLINE=20055634; PubMed=10589717;
RA Halstead J.R., Vercoe P., Gilbert H.J., Hazlewood G.P., Davidson K.;
RT "A family 26 mannannase produced by Clostridium thermocellum as a
RT component of the cellulosome contains a domain which is conserved in
RT mannannases from anaerobic fungi.";
RL Microbiology 145:3101-3108(1999).
DR EMBL: AJ242666; CAB52403.1; -.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000805; Glyco_hydro_26.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF02156; Glyco_hydro_26; 2.
DR PRINTS: PR00739; GLHVDRLASE26.
DR PROSITE: PS00448; CLOS_CELLOSOME_RPT; UNKNOWN_2.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL
FT SIGNAL 1 30 POTENTIAL.
SQ SEQUENCE 589 AA; 66816 MW; 62F45CEB149978F9 CRC64;

```

```

Query Match 10.9%; Score 784; DB 2: Length 589;
Best Local Similarity 35.7%; Pred. No. 1.5e-31;
Matches 204; Conservative 84; Mismatches 185; Indels 98; Gaps 21;

QY 655 RYEAEGFLKGTIVESSGTGYSGAGYVTN-----FHNPGDSLTMITQAPTAGLYNLT 706
Db 37 KYEFEDGHIKGAQIYTDYVQNGEYGEVFDLTGSCSFIAQKGTSTSVNVEVDKELGEIYP 96
QY 707 IGYRSPHD-DKRTNF-SLNGKAFGELLKKKTADPKETSGGKVLNAGANTISFETGNGWY 764
Db 97 ICYVQYDKKKVQYLVNNGVNGQGEISFPFTLKWREISAGIVKLNAGINIELESYGYT 156
QY 765 DIDYVRLEPAADRPDH-AVTKTITNPATVEAKALMNYLVDOYGNKMLSGQEEI----- 817
Db 157 YFDYLVKPADESIVELKVPKLVNPNATVEAKALMSYLVDIYGHILSGQEEICGSHNY 216
QY 818 ----NEIDWLQANVGKKPAIALDLIDYSPSRAEHGLSSTE--AEKAIAMDQ--GGIVTF 870
Db 217 PGSEAEFTYIOEKTGKLPVAVRGDFMNY-----RGNGLMWDQCAERVIEWYKKGGIPTV 272
QY 871 AWHNAPKGLIDTQGEKWRGFEYADSTFTDIEYAMNHPESEYKLLIRIDIVIAQLKKL 930
Db 273 CHWFSFGDI-----GKADNSFTTSTFISRALT-PTGEENIALALLDIDTIAKRLQV 327
QY 931 QDAKVPVLPRLPHEAGKFWGAGKPEPKKLYILMHDRLTNVHKLNNLIWVNS----- 986
Db 328 QDAKVPVLPRLPHEAGKFWGAGKPEPKKLYILMHDRLTNVHKLNNLIWVNS----- 986
QY 987 VAPDWYPGDEYVDILSFSYPOAGDYSPQ-----ISKYEDLVAGKDKKLVAMSENGPIPD 1042
Db 388 TSAAMYPGDGVVDIIGYDKY-NAKDGKPNGSAISSTFYNLVKTNGKLVAMTENDTIPR 446
QY 1043 PDLMKAYQAHWSWATWYGFVDRGCKNSLEHLKKVYNHNPVITLDELPTNLKTYGITE- 1101
Db 447 VSNLVNERAGWLYFCPWGWLITSEQNPNPVDWLVEMVQSDYCIITLDELPTNLKTYGITE- 1101
QY 1102 -----QPSVPGSET-----LNAAGETAKVSLWSANAKSEYKRVSTTENGAFATVA 1149
Db 506 EDSNPDPSPTPTQPKITYGDLNGDKVNSTDLTIMKRYILKNFD-KLAVPEE-----A 558
QY 1150 SDVYGS---SYTDNAV 1162
Db 559 ADLNGDGRINSTDLSI 574

RESULT 10
ID Q9L8L8 PRELIMINARY; PRT: 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BETA-1.4-XYLANASE XYNA PRECURSOR.
GN XYNA.
OS Caldicoccus cellulosovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=20519260; PubMed=11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multidomain 1,4-beta-xylanase from
RT 'Caldicoccus cellulosovorans' and effect of its xylan-binding domain
RT on enzyme activity.";
RL Microbiology 146:2947-2955(2000).
DR EMBL: AF200304; AAF61649.1; -.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR003305; CBD_6.
DR InterPro: IPR001000; Glyco_hydro_10.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF02018; CBD_6; 1.

```

[illegible]

KW Hypothetical protein.

FT NON_TER 1

SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match

Best Local Similarity 56.7%; Pred. No. 1.3e-19;

Matches 97; Conservative 28; Mismatches 41; Indels 5; Gaps 2;

QY 1181 TISNTASAMPKADTQPTTG--LLLOYRTADTKVNDNHLNPOFQIVNKGTSIPINELKI 1238

Db 3 TTPPTSAAP---TTPSAGSLVQYRAADTNAGDNQPKPHFIVNRGTSSVPLSELT 59

QY 1239 RYIITDGTRECFNCDATLSCSLKNGKLVKMEKAATGADYILEVSEFNSDAGVLAPGS 1298

Db 60 RYIYTVGDKPKQVENCDAQVCGSNVRSFYKLSLGTGRTGADYIIITTSAGSLAAGAS 119

QY 1299 TGDITRIHKTWNSYNEDDSYSGTQTSFADHPKVTLYHNGVLVWGTEP 1349

Db 120 SGDIOVRINKNDWTNYNEANDYSYDPTKTSFADWNRVTLNRGQLGVGEP 170

RESULT 14

Q924I1 PRELIMINARY; PRT; 997 AA.

AC Q924I1;

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE CELLULOSE PRECURSOR (EC 3.2.1.4).

GN CELB.

OS Bacillus sp. BP-23.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=89769;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BP-23;

RX MEDLINE=21129642; PubMed=11234960;

RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;

RT "Molecular cloning and characterization of a multidomain endoglucanase

RT from Paenibacillus sp BP-23: evaluation of its performance in pulp

RT refining."

RL Appl. Microbiol. Biotechnol. 55:61-68(2001).

CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

DR EMBL; AJ133614; CAB38941.1; -

DR HSSP; P26221; 10F4.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR003961; FN_III.

DR InterPro; IPR001701; Glyco_hydro_9.

DR Pfam; PF00942; CBD_3; 2.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF00759; Glyco_hydro_9; 1.

DR ProDom; PD001947; CBD_3; 1.

DR SMART; SM00060; FN3; 2.

DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.

DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.

KW Glycosidase; Hydrolase; Signal.

FT SIGNAL 1 34

SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match

Best Local Similarity 6.9%; Score 498; DB 2; Length 997;

Matches 228; Conservative 130; Mismatches 331; Indels 310; Gaps 42;

QY 570 EAGDG-KVHLSWDASSGVVGSVQVATDNGPFTAVASNLTSYTDNTNGTSYYKV 628

Db 90 DAGDHVKFGFPMAASATMLAWSVEYADG----YEAGQLEELK---DNRWATDYPKA 142

QY 629 TAKTNK-----GSSESNILKAVPKMPVNGPARYEAEGLTKGTIVESSGTGYSAGYV 681

Db 143 HTKPNELWGQVAGNTDTHAMWGPAAVVMOMRPS-FKIDASCPCGDLAAETAAALASSY 201

QY 682 TNFHNPGDSLMTTQAAPTAGLYNLITIGYRSPHDDKRTN----- 719

Db 202 FADSDPVYSAKLLQHAKE--LYNFADTYRGKYTDCITDAAAFYNSWTGYEDELAWGAWL 259

QY 720 -FSLNGKAFGELLKTKADPKETSGGKVLNAGANTISFETGW--GWYDIDYVRLEPAAD 776

Db 260 YLATNDNAYLSKAL--SAADRWSSTGG-----SANWPTYWTQGWDSKHY-----GAQ 304

QY 777 RPPHAVTKTLTNPNATVEAKALNLYVDQYKGNLSCQEEINEIDWLOANVGKKPATAAL 836

Db 305 ILLARIITSNLMPDPT-----KFIQSTERNLDY--WTVGTVNGR----- 341

QY 837 DLIDYSPRAEHLGSLSTEAEKAIAMWDKGGIVTFAMHWNAPKGLIDTQGWK-----WRG 891

Db 342 --VKYTPG-----GLAWLDQWGLRYA---ANAAFISFVYSWVSDPVKKS 382

QY 892 FYADSTTFDIEYAM-NHPESEDYKLLIRDDIVTAGQLKKLQDAKVPVLFPLHE-ABGKW 949

Db 383 RYQNFATSQINYLGDNPROSSY-----VVGQNSPQ-----HPHRTAHSSW 426

QY 950 -----FWMGAK--GPEPVKKLYILMHDRLTN-----VHKLNLIW 982

Db 427 MNNEIDIPANHRHILYGAUVGPNASDQYTDIDGYVSNEVATYNAGFTGALAKMNLLYG 486

QY 983 -----VWNSVADPWYPGDE-----YVDILSF-----DSYP-QAGDYSPOIS-K 1018

Db 487 QNHQPLANFPAPE-VKGDEYFVEAAVRSSGNTETIRALLNNRSGWPARGMD--QLSPK 542

QY 1019 Y----EDLVALGKD---KKLVAMSENGPIPDPLMKAYQAHWSWEPATWYGDVVRGKQN 1070

Db 543 YFIDLSEVYAGRTVSDVQTVTSSEGATVSPVWVDAAKRIYAITANFSNTKIYFGE 602

QY 1071 SLEHLKKVY-----NHPNVITLDELPTN-LKTY-----GIT-- 1100

Db 603 N--YRKEVQFRITGPGQAWNPANDPSYQNLTTGPNPKSVNIPYDAGVKVSGOEPGTPV 660

QY 1101 EQSPVSGSFTLNAGETAKVSLSWTASANKSYEVKRSRTTENGAFATVADSVGYSSYTD 1160

Db 661 AVFAAPAG--VQAVAGNSQVALNWSAGAVSYTVKRAEVSGGPTTVAAGVNGLYTNT 718

QY 1161 AVTADTTYQYQVAKNDAGOT-LSNTASAMPKADTQOP----- 1197

Db 719 GLINGKTYIYVTVAVNSAGESPASVQVSGMPQAAATTVPGAVTLSTAGNQNMLSWTAS 778

QY 1198 ----- 1197

Db 779 GASSYTVQRAVAGGTVTDVATGLAVLNNDTTALNGTSYSYRVAAVNASQOTLSNVMLT 838

QY 1198 -----TTGLL-LQYRTADTKVNDNHLNPOFQIVNKGTSIPINELKIRYYTIDGRBO 1250

Db 839 PSGPPVSTGTLEQYRSGSGNSNAVTPQNLKNTGTQALDLSLKIRIYFTKDGTEEL 898

QY 1251 TENCYATLSCSLKNGKLVKMEKAATGADYILEVSEFNSDAGVLAPGSGTGTIOTRIHKTD 1310

Db 899 SFWCDDAQGSANVQGMFVAVNPAKGTADTYVEISFTSAGSLAAGAETGVIOTRFSKN 958

QY 1311 WSNYNESDDYSYKGTQTSFADHPKVTLYHNGVLVWGTEP 1349

Db 959 WSAFDQSDNDYSYDASKATAFAANKVYAYQGNQVWGLEP 997

RESULT 15

Q9ADL2

ID Q9ADL2

AC Q9ADL2;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)

DE BETA-MANNANASE.

OS Polyangium cellulosum.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 18:45:14 ; Search time 18262.7 seconds
(without-alignments)
4647.619 Million cell updates/sec

Title: US-09-784-554B-3
Perfect score: 4056
Sequence: 1 atgaagcgcaaaatagtag.....ccgaaccgacagctaattaa.4056

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	630.2	15.5	2977	6	A28170	A28170 B.lautus
2	629.4	15.5	1854	6	AR012245	AR012245 Sequence
3	629.4	15.5	1854	6	IL4943	IL4943 Sequence 28
4	629.4	15.5	1854	6	I73663	I73663 Sequence 28
5	627	15.5	2831	1	BACCELAA	M76588 Bacillus la
6	523.2	12.9	6465	1	D83704	D83704 Clostridium
7	439.6	10.8	13603	1	AE007607	AE007607 Clostridi
8	417.4	10.3	6005	1	AF078038S5	AF078042 Caldicell
9	417.4	10.3	6416	6	E35100	E35100 Truncated c
10	391.2	9.6	4977	1	CDCMANA	L01257 Caldocellum
11	327	8.1	1433	1	AF013989	AF013989 Dictyoglo
12	251	6.2	3900	1	U39812	U39812 Caldicellul
13	193.6	4.8	2492	1	RFU08621	U08621 Ruminococu
14	189.4	4.7	4831	1	RMRMWAG	X90947 R.marinus m
15	172.6	4.3	2692	1	RFL298117	AJ298117 Ruminococ
16	172	4.2	4567	1	AF163837	AF163837 Caldibaci
17	170.2	4.2	3237	1	AF200304	AF200304 Caldibaci
18	143.2	3.5	4161	1	BSP133614	AJ133614 Bacillus
19	137.6	3.4	2201	1	AB044406	AB044406 Clostridi
20	133.6	3.3	1855	1	AF078038S2	AF078039 Caldicell
21	114.8	2.8	2020	8	PSEI4BMAN	X97408 Piromyces s
22	111.2	2.7	2336	8	PSMANAGEN	X91857 Piromyces s
23	107.4	2.6	1862	8	PSMANC	X97520 Piromyces s
24	105.6	2.6	2015	1	CTH242666	AJ242666 Clostridi
25	99	2.4	3180	1	AY039744	AY039744 Bacillus
26	90.6	2.2	1920	1	BACGLUB	M16185 B.subtilis
27	90	2.2	11707	1	AF078737	AF078737 Caldicell
28	90	2.2	11707	6	E35099	E35099 Truncated c
29	87.8	2.2	5439	1	CDCCELA	L32742 Caldocellum
30	87.6	2.2	3262	1	AF078038S1	AF078038 Caldicell
31	87	2.1	2175	1	BACCELD	M28332 B.subtilis
32	87	2.1	2175	6	E05425	E05425 DNA sequenc
33	86.2	2.1	5284	1	CSUL6308	U16308 Caldocellum
34	85.8	2.1	1800	1	BSEGLSG	Z29076 B.subtilis
35	85.8	2.1	1928	1	BACMCASE	D01057 B.subtilis
36	85.8	2.1	2314	1	BS14GLUC	X67044 B.subtilis
37	85.8	2.1	2435	1	BSBGLUC2	X04689 Bacillus su
38	85.8	2.1	26170	1	BC170DRGR	Z73234 B.subtilis
39	85.8	2.1	233780	1	BSUB0010	Z99113 Bacillus su
40	85.6	2.1	2587	1	AB016164	AB016164 Bacillus
41	85.2	2.1	2029	6	E35142	E35142 Truncated c
42	85.2	2.1	2029	6	E35143	E35143 Truncated c
43	85.2	2.1	5437	1	CASR69XYN2	AF036924 Caldicell
44	84.6	2.1	4241	1	CSCELB	X13602 Caldocellum
45	84.2	2.1	1523	1	AF355629	AF355629 Bacillus

ALIGNMENTS

RESULT 1
A28170
LOCUS
DEFINITION B.lautus strain NCIM 40250 endocellulase gene 1 (Endo1).
ACCESSION A28170
VERSION A28170.1 GI:905293
KEYWORDS
SOURCE
ORGANISM
Paenibacillus lautus.
Paenibacillus lautus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Paenibacillus.
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1 (bases 1 to 2977)
/organism="Paenibacillus lautus"
/db_xref="taxon:1401"
677..2779
/codon_start=1
/transl_table=11
/product="endocellulase 1"

/organism="unknown"	
BASE COUNT	
452 a	506 c
503 g	393 t

Db 201 GCTGACGGTTTACACTGGGAGACACACGATCCAGCGCCGGAAGGACTGGCTTCATTA 260
Qy 312 cagcgatagctattatgcgacacgcggtattgacaaaagccgaatgtgaaagccagg 371
Db 261 CAGCGATGATTCTCTCGCGCAACGGTGGTGTTCAGACACGACGCTGCACAGCGCGG 320
Qy 372 tgcggtggaacacccgttctacgataaactcgtggaacgagcaggaacacattcttagtca 431
Db 321 GCGGCTGTTTACCGCTTTTTCACGATAAATCTTTGGAGAAATGGAGCTTACCTCCATTTGAAC 380
Qy 432 actgcacatggtccggttatgtgccaagatggaacggaagtgtgcagaaagcaaaa 491
Db 381 GCTGCAATAGCGGGTATGTGTCGCCGGATAAGAACGGTCCAGTGTACCGAGAGTGAGAC 440
Qy 492 ggtcccttcgctcgttggaatgaggtctaaacgctaaaaatgcgcgttttcaattgca 551
Db 441 GCGTCCGTCACCGGTTGGGATAAGTTCGAGTTTGGCCAAAATGCGCGTCTCCCTTCA 500
Qy 552 gctgatctgaagaacacatcaggtttatgcggtatggaatcgtcaacttttagtgaaaa 611
Db 501 GCCTGATCTGAACGACGCGCAAGTGTATATGGATGAAGAAATTAACCTTCCTGCTCAACCG 560
Qy 612 gtacgggtgtcttcaaaaaacgcgcgtgaaagatactcgtcgaacatgaacccgc 671
Db 561 GTATGGAACGCTCAACGCTCAACGGGCATCAAGCGTATTCGCTGGATTAACGAGCGCGC 620
Qy 672 tctctggtcgcatacgcacgcgcgcattcattggtgaaaggtcgcgagcgaagagttggt 731
Db 621 GCTGTGCTGAGACGATCCAGGATTCATCCGAGCAGTTCACAGCGGCACAGCTCGT 680
Qy 732 agaccggtcggtaagttatccaaagccgctaagggcgttgacgcgggtgcggaatttt 791
Db 681 CGCTAAGACATCGACTGTGTCAAAGGCGGTGAAGAACGTGCTGATCCGCTGCTGCTGCTGCT 740
Qy 792 tggcccggtttctacggttttgcgcctataaagatcttcaaacgcacacctgattgaa 851
Db 741 CGGTCTCTGCTTACCGGTTTCGCGGCAATTTGCTCTCAGAGCAGCAGCGGATTTGGCC 800
Qy 852 ctctgtaaaaaggaactcacagctggttctgtggaactattactcgtcgaacatgcgcctcag 911
Db 801 GAGTTTCAAGGCAACTACAGCTGGTGTATTCGATTACTATCTGATCAGATGAAGATGC 860
Qy 912 ctgcgaagccgaaggaagatgctggtgtctctcgtatgctacactggtatcctgaagc 971
Db 861 TCATACGACGAAAGGCAAAAGATTGCTCGATGCTGCGGATCCGCTGCTGCTGCTGCTGCTGCT 920
Qy 972 gatggcgaggagacacgaattacaaatga---ggtaggcaacgacgaacgaagaaac 1028
Db 921 ACAGGCGGAGGCGCAGGCAATCTCTTTGCGGGGGCGGCAATATCGATACGCGAAGGC 980
Qy 1029 cagaatgcaagcgcctcgttctgttgggatacgcacactcaagaagagatagctggatgc 1088
Db 981 TCGCGTACAAAGCGCAAGATCGTATGGGATCCGGTTCACGAGAAAGACAGCTGGATCGG 1040
Qy 1089 tcaatgaaacagtgaaattcttcttactgctcgtcgtatgaagcagctcgttggaata 1148
Db 1041 CACATGTTTTCAGCTACTTTGCTTAAATTCGGAAGCTGCAATCTTCGATTTCAGACGTA 1100
Qy 1149 ttaccggggaacaaagctgcttctgagtagatagctggtggtgcaaaaacagctttatggc 1208
Db 1101 TTATCCGGGTACCAAGCTGCGATCAGAGTTACGTTACGTCGGGAGAGCAATCACATTTC 1160
Qy 1209 cggcggtatcgtatggtcgtatgctggtggaatcttgggcaaaaacagctttatggc 1268
Db 1161 GGGAGGCATAGTACCGCGGACGCGCTCGGCATTTTGGAAAATATGGCGTTTATGCCGC 1220
Qy 1269 aaactaatgaaagtaagatggttccaaacactacgttagtgcgtctacacgttta 1328
Db 1221 GAATTAAGTGGCAG---ACGAGAGCAATACCGATTATACCGGCTGCTTACAGCTGTA 1277
Qy 1329 ccgcaattatgacggaagagcttacttctcgtgatatcagcgttcatcgcgcaacgctc 1388

Db 1278 TCACAACACTACGAGCGCAATAAATTCGGGGTTCGGCTCGATCAAGTGGAGCGCGCTACGTC 1337
Qy 1389 ggaattgttaatagctcgtgctcatcttcctgtaaacgagtgcatcctacaagaactgca 1448
Db 1338 CGATACGGAGACAGCTCGGTATACGCTTCGGTAACTGACGAGGAGAAATTCGGAATCCA 1397
Qy 1449 cctcgttgtcaataaagcatggaacagtgacagtcattcgacgcccaatttgatctttccgg 1508
Db 1398 CCTGATCGTCTGAATAAATAATTCGACGATCCGATCAACGCTACTTTCAGCTGCTCTGG 1457
Qy 1509 ccagacgacttcaggttcggttaaaagtatggttttcgacaaaaataagtcgcgaattaa 1568
Db 1458 TGATAAAACCTACACATCCGGGAGAGTATGGGGCTTCGACCAACCGGATCCGACATTAC 1517
Qy 1569 ggaagcagcgcaatcaccaaatcttcagcgaacggttttactctacagtagtaccgcttt 1628
Db 1518 GGAACAAGCAGCTATTAACGAATATTAAACAATCAATTCACGTATACGCTTCCTCCATT 1577
Qy 1629 gacggtttatcacatcgtgttgactgcgggcaatgatacacctgt 1673
Db 1578 GTCGGCTTACCACATGTTCTGAAAGCGGATAGCACCGAACCGGT 1622

RESULT 4
LOCUS I73663 1854 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 28 from patent US 5686593.
ACCESSION I73663
VERSION I73663.1 GI:3009804
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Woldike, H. Fabricius, Hagen, F., Hjort, C., Mailand, and Hastrup, S.
TITLE Enzyme capable of degrading cellulose or hemicellulose
JOURNAL Patent: US 5686593-A 28 11-NOV-1997;
FEATURES
Location/Qualifiers
1..1854
/organism="unknown"
BASE COUNT 452 a 506 c 503 g 393 t
ORIGIN

Query Match 15.5%; Score 629.4; DB 6; Length 1854;
Best Local Similarity 63.8%; Pred. No. 9.1e-167;
Matches 986; Conservative 1; Mismatches 552; Indels 6; Gaps 2;

Qy 132 tgttaccattaaagtcgatacatccaaggatcgtaagcctattagcccttattattacgg 191
Db 81 TGTCACCTTTCAGATTAATACGACGTCGGAACGTCGACGATCAGCCCAATATTTACGG 140
Qy 192 tacgaatcagagttgagcgatgagaatctgactccagacgactggtgcaatcg 251
Db 141 AACCAATCAGGATCTGACGGGACGCGGAACACTGGTTCATCCCGCAGGCTCGAGGCAACCG 200
Qy 252 aatgaccggatataactgggaaaaaataatgtcccaatcaggaagcagactggtgagtc 311
Db 201 GCTGACGGTTACACTGGGAGACACACGATCCAGCGCCGGAAGGGAATGGCTTCATTA 260
Qy 312 cagcgatagctattatgcgacacgcggtattgacaaaagccgaatgtgaaagccagg 371
Db 261 CAGCGATGATTCTCTCGCGCAACGGTGGTGTTCAGACACGACGCTGCACAGCGCGG 320
Qy 372 tgcggtggaacacccgttctacgataaactcgtggaacggaagtgtgcaggaagcaaaa 431
Db 321 GCGGCTGTTTACCGCTTTTTCACGATAAATCTTTGGAGAAATGGAGCTTACCTCCATTTGAAC 380
Qy 432 actgcacatggtccggttatgtgccaagatggaacggaagtgtgcagaaagcaaaa 491
Db 381 GCTGCAATAGCGGGTATGTGTCGCCGGATAAGAACGGTCCAGTGTACCGAGAGTGAGAC 440
Qy 492 ggtcccttcgctcgttggaatgaggtctaaacgctaaaaatgcgcgttttcaattgca 551

120 aacqqcaaaqactqttaccattaaagtccgtacatccaaagcatcgtaacccctattacccc 170

Db	1643	GATTGACGCTATTATTCGGGTACGAAGCTGGCATACAGAGTCCAGCTACGCGGAGA	1702
QY	1197	aaatgatattccggcggtatcgctatgcccgaatgctcgggcatcttgggcaaaaacga	1256
Db	1703	CAATCACATTTCGGAGGCATAGCTACCGGACGCGCTCGGCATTTTGGAAATATGG	1762
QY	1257	cgtttatatggcaactactggaagttaaaagatggtcgcaaaactactgttagtcgcgc	1316
Db	1763	CGTTTATGCGCGAATTACTGCGACAGGAGGACAAATACCGATTATACC---AGCGGTGC	1819
QY	1317	ttacaagctttaccgcaattatgacggaaaaagctctacttttcgttgatcatcagcgctca	1376
Db	1820	TTACAAGCTGTATCGCAACTACGACCGGCAATAAATCGGGGTTCGGCTCGATCAAGTGA	1879
QY	1377	tgcgcaaacgctcggtattgttaatagctcgtggtcatgttccgtaacgtagatcatccta	1436
Db	1880	CGCCGCTACGTCGATACGGAGAACAGCTCGGTATAGCTTCGGTAACTACGAGGAGAA	1939
QY	1437	caaaagaactgcacctcgtgtgcatgaaataaagcatgagcagtgcatctgcagccccaatt	1496
Db	1940	TTCCGAACCTCCACCTGATCGTGCTGAATAAANAATTTGACAGATCCGATCAACGCTACTTT	1999
QY	1497	tgaattttccggcgagacgaattcaagggttcggttaagatgtaggggtttcgacaaaaatag	1556
Db	2000	CCAGCTGTCTGGTGATATAAACCTACATCTCGGGAGAGTAGTGGGGCTTCGACCAACCGG	2059
QY	1557	ctcgcaaatlaaggaagcagcgccaatcaccaaatctcaggcaaacggtttacctatcac	1616
Db	2060	ATCCGACATTACGGAACAAGCAGCTATAACGAATATTTAACACACANTCAATTACGTATAC	2119
QY	1617	agtaacgcctttgacgggttatcacatcggtgttgactgcgcggaatgatatacacctgt	1673
Db	2120	GCCTCTCCATTGTGGGTTTACCACATTGTCTGAAAGCGGATAGACACCGAACCGGT	2176

RESULT	6
D83704	
LOCUS	D83704
DEFINITION	Clostridium thermocellum DNA for endoglucanase, complete cds.
	5465 bp DNA linear BCT 06-FEB-1999

LOCUS	D83704	6465 bp	DNA linear	BCT 06-FEB-1999
LOCUS	D83704	6465 bp	DNA linear	BCT 06-FEB-1999
DEFINITION	Clostridium thermocellum DNA for endoglucanase, complete cds.			

RECESSION D83704
VERSION D83704.1 GI:1663518

KEYWORDS endoglucanase.

SOURCE Clostridium thermocellum (strain:F1) DNA.

[illegible]

Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium

REFERENCE
1 (bases 1 to 6465)

AUTHORS Sakka, K.

TITLE Direct Submission

JOURNAL Submitted (25-FEB-1996) Kazuo Sakka, Mie University, Faculty of

Bioresources; 1515 Kamihamacho, Tsu, Mie 514-8507, Japan

REFERENCE 2 (pages 1 to 645)
(E-mail: sakka@bio.mie-u.ac.jp, Tel: 0592-31-9621, Fax: 0592-31-9634)

REFERENCE
AUTHORS
Z (Dases I LO 6463)
Sakka K

JOURNAL
UNPUBLISHED (1996)

REFERENCE

AUTHORS Ahsan, M.M., Kimura, T., Karita, S., Sakka, K. and Ohmiya, K.

TITLE Cloning, DNA sequencing, and expression of the gene encoding

Clostridium thermocellum cellulase CelJ, the largest catalytic

JOURNAL OF BACTERIOLOGY, Vol. 179, No. 5, 1995, p. 1790-1795.
Copyright © 1995, American Society for Microbiology

J. BACTERIOL. 178 (19), 3732-3740 (1996)
MEDLINE 96422002

Sequence updated (30-Oct-1996) by: Kazuo Sakka

FEATURES

```
source
1. .6465
```

/organism="Clostridium thermocellum"

```

/strain="F1"
/strain="F1"

```

505 5400
/db_xref="taxon:1515"

```
gene
393.3400
/anno="celr"
```

CDS	595	5400
595	5400	

```
/gene="celJ"
```

```

/codon_start=1

```

Matches 893; Conservative 1; Mismatches 642; Indels 51; Gaps 4;

QY 132 tgttacattaaatcagatccatccaaagatcgtaagcctattagcccttattattacgg 191
 Db 4312 TGTAAATATCGATCGTCCAGGGAAGAAACAAAGATAGCCCGTATATTTATGG 4371
 QY 192 tacgaatcagagtggtgagcgatgagaaatcctgactgcccagacgacttgggtgcaatcg 251
 Db 4372 AGCAATCAGGATATCAGGGTGTGTACCC---CTGCAAGACGACTTGGTGGGAACAG 4428
 QY 252 aatgacggatataactgggaaacaaatattgccaatgcaaggagcactggatgcagtc 311
 Db 4429 ATTGACGGGTACAAATGGGGAACAAATATGTCCTCAATGACAGGAGTACTGCTATCATTC 4488
 QY 312 cagcgatagctattatgcaacaacgcggatgacaaaagcgaatgtgaaaagccagg 371
 Db 4489 AAGCGATGATTATATGTTTATATATGTTGATTAACAGGGAATGATAGAACGTTCCAGC 4548
 QY 372 tgcggtgcaacctcgcttccagcatcaatcgctgaagcagggcacatattcttttagtcac 431
 Db 4549 AGCTGTTGTAAGCAATTTACAGAGCAGTCAATAAAGCAAAATGCATATTCAGCCATCAC 4608
 QY 432 actccggtgcccgttatgtgcccagagtggaacggaagtgtgcaggaagcgaacaa 491
 Db 4609 ATTACAGATGTTAGTTATGTTGCAAGAGTGGGAATGTTACAGTGAGCGAGTCAGAGAC 4668
 QY 492 ggcctctccgctcgggtggaatgaggtcgtaaacgcgttaaaatgcgcggtttcaattgca 551
 Db 4669 AGCTCCGTCGGGAGATGGGTGAGGTCAGGTTTAAAGAAAGATGGTGCACGTGCTATTGCA 4728
 QY 552 gctgatctgaagacaactcaggtttatgcggtgaatctgcaacttttttagtgaacaa 611
 Db 4729 GCCTGACGTGAATGATTAATGATATATATGATGAGTATTTAACTATCTGATTAATAA 4788
 QY 612 gtaogcgttcttccaaacaaacggtggaaggtataactgcgtcgaatgaacccgc 671
 Db 4789 GTATGTCGATCATCGTCTGCAAGCGGAATTAAGAGGATATATACCTTGACACAGCGCGGA 4848
 QY 672 tctctggtcgatcacgactcgcgcattcatgtggaagcgtcggaagcgttggtggt 731
 Db 4849 CTTATGTTTACTACTCTCCGGAATTCATCCACAGNAGGTAACCTGCAGTGAATGAT 4908
 QY 732 agacggtcggtgaattttatccaaacggtcgaagcgttgacggtggtggtggtggtggt 791
 Db 4909 AAATAATCGTGGAGCTGGCGAAAGTAATAAGACACACTTGCATCCAGATGAGAAATTT 4968
 QY 792 tgggcccgttcttaacggttttgcgcctataaagatcttcaaacctgcacactgattgaa 851
 Db 4969 TGGACCTGCATCGTATGTTTGTGGGATATTTAACATTTGCAGGATGCACCTGACTGGA 5028
 QY 852 ctctgtaaaagcactacagctggttcgtgactattacctgcgtcaaatgcgcctcag 911
 Db 5029 TCAGGTTAAAGGAATACACATGTTTGTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTT 5088
 QY 912 ctgcgaagccgaagcaagattgctggtatgtcttcgatgtacactggttatcctgaagc 971
 Db 5089 ATCGGATAGTTTGTGGGAARAGGTTATTGGATGTACTTGCATACACTGTTACCCGGAGGC 5148
 QY 972 gatggcgaggagcatcacgaattacaaatgaggtaggcaaacac-----gaaacgaa 1022
 Db 5149 GCAGGTTGGCGGTGGCGGAATATGCTTTGACGGTGAATAATAGTACTTCAAGGGATGTGGC 5208
 QY 1023 gaaagccgaatgcaagcgcctcgtacttctgtgggtatccacactacaag----- 1071
 Db 5209 AATAGCGAGGATGACAGGCCAGGACCGTATGGATCCGACATATAAAACCCACCCAGAA 5268
 QY 1072 -----gaagatagctggtatgctcaatggaacagctgaattcttgc 1112
 Db 5269 AGGTGATGATAACAGCGGGAGAAATAGCTGGATAAACCAATGTTTCCAGGATATCTTCC 5328
 QY 1113 ttactgctcgtattaaagcagctcgttgataagttattaccggtggaacaaagctggttt 1172
 Db 5329 ACTGTTCCCAATATAAAGGCGAGATATAGACAAGTATTATCTCTGTGTAACCAACTGCTTAT 5388

QY 1173 gactagtagctatggtggtgcaaaatgatatttccggtggtatcgtatggtcgtatgt 1232
 Db 5389 AACTGAGTTTGTATTGAGGGAAGACCATATATCGGGAGGAATAGCTTTAGCAGATGT 5448
 QY 1233 gctgggcatcttgggcaaaacagcgtttatatgcaaaactactggaagttaaaagatgg 1292
 Db 5449 GTTAGGATATTCGGCAAGTATGGAGTATACATGGCAGCAAGATG-----GGGAGA 5499
 QY 1293 tgcacaacaactacgttagtgccgcttacaagctttaccgcaattatgacggaacaaagctc 1352
 Db 5500 TTCCGGGAGCTATGACAGCGCGGTACAACTTTTATCTCAACTATGATGGGAAAGGTTTC 5559
 QY 1353 tacttccgtgatcatcagcgttcacgcgaacgctcggtatattgttaatagctcgtgtga 1412
 Db 5560 GAGATACGGTTTCACTGCTGAGCGCTGAGACAACTGACGTTGAGAACATGCCGGTATA 5619
 QY 1413 tgcctccgtaacggatgcatcctcacaagaactgcacacctcgttgcgtatgaaataaagcat 1472
 Db 5620 TGCTTCAATTGAGGGAGAAAGATGATTCGACTGTGCATATATATTAATTAACAGGAATTA 5679
 QY 1473 ggcagctgcatcgcagcccaatttgatcttccggtgagagcacttaccgttccggttaa 1532
 Db 5680 TGACAGGAACATGAAGGCGAGATAAAGATGAATAATACAGGGTATACACAGGTGGAGA 5739
 QY 1533 agtatgggttttcgcaaaaatagctcgcgaatgaaggaagcagcgcccaatcacscaat 1592
 Db 5740 GATATACGATTTGACAGTACAAGCTCTCAGATCAGAAAGATGGAGTCTCAGTAATAT 5799
 QY 1593 ttcaggcaacccggtttacatctacagaccgcttgcgcttgcagcgttatcacatcgtgtgac 1652
 Db 5800 ACAAACAACACACATCCATAGAGTTCGAAATCTGACGGTATACCATATTGTTTAAAC 5859
 QY 1653 tgcggcgaatgatacacctcgtgagaaa 1679
 Db 5860 TTCTTCAAGTAGATTAAAGAAATAAAA 5886

RESULT 9
 E35100 6416 bp DNA linear PAT 07-FEB-2001
 LOCUS Truncated cellulase composition.
 DEFINITION
 E35100
 E35100.1 GI:13018925
 VERSION
 KEYWORDS JP 1999221086-A/2.
 SOURCE
 unidentifed.
 ORGANISM
 unclassified.
 REFERENCE
 1 (bases 1 to 6416)
 AUTHORS
 Daiji,A.P.L., RoyL,M.D.G.G., Farinton, Moreland,D.G.H.H. and
 Daian,P.W.
 TITLE
 Truncated cellulase composition
 JOURNAL
 Patent: JP 1999221086-A. 2 17-AUG-1999;
 CLARIANT INTERNATIONAL LTD
 COMMENT
 OS Unknown
 PN JP 1999221086-A/2
 PD 17-AUG-1999
 PF 21-SEP-1998 JP 1998283606
 PR 19-SEP-1997 US 08/932571
 PI PALU ANDERSON, PETER L BAGUKUISUTO, ROY M DANIEL, PI
 GURAHAM K FARINTON,
 PI MORELAND DAVID GIBUSU, HYU MORGAN, DAIAN PURATONORIUS WILLIAM
 PC C12N15/09, C11D3/386, C12N1/21, C12N9/42//C12N1/21, C12R1/19), PC
 (C12N9/42, C12R1/19), C12N15/00
 CC
 FH Key Location/Qualifiers
 FT source 1. .6416
 FT /Organism= 'Unknown'.
 FEATURES
 source
 1. .6416
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 2068 a 1082 c 1689 g 1541 t 36 others

ORIGIN

Query Match	10.3%	Score 417.4	DB 6	Length 6416
Best Local Similarity	56.3%	Pred. No. 1.3e-106		
Matches 893	Conservative 1	Mismatches 642	Indels	Gaps
QY 132	tggtaccattaaagtcgatacatccaaggatcgtaagcctattagcccttattatcagg	191		
DB 4329	TGTTAAATATCGATCGATAGCTCCAGGGGAAGAACAAAGATANGCCGTATATTATGG	4388		
QY 192	tacgaatcaggagttggcaggcgatgagaatctgactgcgcagacaacttgggtggcaatcg	251		
DB 4389	AGCAAATCAGGATATCCAGGGTGTGTTCACC---	4445		
QY 252	aatgaccggatataacttgggaaaaaatatgtccaatcgagggaagcgactgatcagctc	311		
DB 4446	ATTGACGGGTTACAAATTGGGAGAACAAATATGTCCAATGACGGAGTGACTGATCATTC	4505		
QY 312	cagcgatagctattatgacaaacgccggattgacaaagccgdatgtgaaaaaccagg	371		
DB 4506	AAGCGATGATTATATGTGTTATATTATGTTGGTATTAACAGGGAATGATAAGACGTTCCAGC	4565		
QY 372	tgcggtggcaacctcgtttcacgatcaatcgcgtgaagcagggcacatatatttttagtca	431		
DB 4566	AGCTGTTGTGAAGCAAAATTCACGACGAGTCATAAAGCAAAATGATATTCAGCGATCAG	4625		
QY 432	actgcgatgcccgggttatgtgccaagdatggaacggaagtgtcgagaaagcgaaaa	491		
DB 4626	ATTACAGATGTTAGTGTATTGTGGCAAGGATGGGAATGTGTACAGTGAGCGAGTCAGAC	4685		
QY 492	gpcctcttcgcctcgggtggaataggtcgtaaacgcctaaaaatgcgcgtttcatttgc	551		
DB 4686	AGCTCCGTCGCCGAGATGGGCTCAGGTCAAGTTTAAAAAAGATGTGTGCATGTCATTGCA	4745		
QY 552	gcctgatctgaagacaactcaggtttatcgcatgaattcgcgaacttttagtgaaaa	611		
DB 4746	GCCTGACGTGAATGATAACTATGTTATATGATGATGAGTTTATTAACTCTGTTAATAAA	4805		
QY 612	gtacggcggttgcttcaacaaaaacgggcgtgaaagatactcgcctgcagaaatgaaccgc	671		
DB 4806	GTATGGTTCGATCATCGTCTGCAACGGGAATTAAGGATATATACTGTGACACGAGCCGA	4865		
QY 672	tctctgtgcatacgcataccgcgaattcatggtgaaaaggtcgagagcgaaaagattggt	731		
DB 4866	CTTATGTTTACTACTATCCCGGAATTCATCCACAGAAGGTAACCTGCAGTGAATTTGAT	4925		
QY 732	agacggctggaagtattccaaaacgcctgaagcgggttgacgcgggtgcggaaatttt	791		
DB 4926	AAATAAATCGTGGAGCTGGCAAGTAATAAAGACACTTGATCCAGATGCAGAAATTTT	4985		
QY 792	tgggcccgtctcttaacggttttgccctataaagatacttcaaaactgcacctgattggaa	851		
DB 4986	TGGACCTGTCATCATGTTGTTTGTGGATATTTAACTATGTCAGGATGCACCTGACTGGA	5045		
QY 852	ctctgtaaaaaggaacatacagctggttcgttggaattattacactgcatacaaatgcgcctcag	911		
DB 5046	TCAGGTTAAAGGAATACAGATGTTTTTGAGCTGGTACCCTTGAGCAGATGAAGAAGC	5105		
QY 912	ctcgcgaagccgaaggaagattgctggatgctcttgatgtacactgggtatcctgaagc	971		
DB 5106	ATCCGATAGTTTTTGGGAARAGTTATTGGATGCTACTTGACATACACTGGTACCCGAGGC	5165		
QY 972	gatggggcgaagcatcacgaattacaaaatgaagtgaggcaacgac-----gaaacgaa	1022		
DB 5166	GCAGGTTGGCGGTGTCGGAATATGCTTTTGACGGTGAATAGTACTTCAAGGATGTGGC	5225		
QY 1023	gaaagccagaaatgcaagcgctcgtactttgtggatccgacctacaag-----	1071		
DB 5226	AATAGCGAGGATGCAGGCACCGAGAACGCTATGGGATCCGACATATAAACCACCCAGAA	5285		
QY 1072	-----gaagatagctgcatcgtcctaattgaaacagtgaaattcttgc	1112		

Db	5286	AGGTCAGATAACAGCGGGAGAAAATAGCTGGATAAACCAATGGTTTTCCAGAGTATCTTCC	5345
Qy	1113	tttactgcctcgattaaacgagtcggtgataagattaccgcggaacaacgctgcttt	1172
Db	5346	ACTGCTTCCCAATATAAAGCGCAGATATAGACAAGTATTATCTCGTGATACCAACTTGCCTAT	5405
Qy	1173	gactgagtagatagtggtggcgaaaaatgattttccggcggtatcgctgatacgccgatgt	1232
Db	5406	AACTGAGTTTGATTTATGGAGGGAAGGACCATATATCGGAGGAGTAAGCTTTAGCAGATGT	5465
Qy	1233	gctgggcatcttggcgcaaaaaacgactttatatggcaaaactactggaagttaaagatgg	1292
Db	5466	GTTAGGGATATTTCGCGAAGATGGAGTATACATGGCAGCAAGATG-----GGGAGA	5516
Qy	1293	tgcacaactacgttagtcgcgtctcaagctttaccgcaattatgacggaaaaagctc	1352
Db	5517	TTGCGGGAGCTATGCACAGCGCGCTACAACTTTATCTCAACTATGATGGGAAAGGTTTC	5576
Qy	1353	tacttcggtgatatacagcgttcacgcgcaaacgtcggtatgtttaatagctcggtgca	1412
Db	5577	GAGATACGGTTCAACGTGTGTGAGCGCTGAGACAACACTGACGTTTGAGAACATGCCGGTATA	5636
Qy	1413	tgcttcgtaacgagatgcatctcaaaagaactgcacctcgttgatgaataaaagcat	1472
Db	5637	TGCTTCAATTGAGGGAGAGAATGATTGACACTGTCGATATATATTAATTAAACAGGAATTA	5696
Qy	1473	ggacagtcattcagccccaatttgatcttccggcgagacgacttcaggttcgcgtaa	1532
Db	5697	TGACAGGAACTGAGGCGAGAGATAAGATGAATTAANTACCAGGGTATACACAGGTGGAGA	5756
Qy	1533	agtagggggttcgacaaaaatagctgcgaataaaggaaagcagcgccaatcacscaat	1592
Db	5757	GATATACGGAATTGACAGTACAAGCTCTCAGATCAGGAAGATGGGATGCTCAGTAATAT	5816
Qy	1593	ttcggcaacggyttacctatagtcaccgcctttgacggccttatcacatcggttgac	1652
Db	5817	ACAAAACAACAACATCACCATAGAAGTTCCAAATCTGACGCTATACCATATTTGTTTAAAC	5876
Qy	1653	tgccggcaatgatacacacctgtagaaa	1679
Db	5877	TTCTTCAAGTAGATATAAAGAATAAAA	5903
RESULT	10		
LOCUS	CDCMANA	4977 bp	linear BCT 26-APR-1993
DEFINITION	Caldocellum saccharolyticum beta-mannanase/endoglucanase (mana)		
ACCESSION	gene, complete cds.		
VERSION	L01257.1 GI:144290		
KEYWORDS	beta-mannanase; endo-1,4 beta-mannanase; endoglucanase; mana gene.		
SOURCE	Caldocellum saccharolyticum (library: lambda NZP2) DNA.		
ORGANISM	Caldicellulosiruptor saccharolyticus		
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group;		
AUTHORS	Syntrophomonadaceae; Caldicellulosiruptor.		
TITLE	1 (bases 1 to 4977)		
JOURNAL	Gibbs, M.D., Saul, D.J., Luthi, E. and Bergquist, P.L.		
MEDLINE	The beta-mannanase from Caldocellum saccharolyticum is part of a		
FEATURES	multidomain enzyme		
Source	Appl. Environ. Microbiol. 58, 3864-3867 (1992)		
gene	93119139 Location/Qualifiers		
CDS	1..4977		
	/organism="Caldicellulosiruptor saccharolyticus"		
	/db_xref="taxon:44001"		
	841..4836		
	/gene="mana"		
	841..4836		
	/gene="mana"		
	/EC_number="3.2.1.78"		
	/note="endoglucanase EC_Number= 3.2.1.4"		
	/codon_start=1		
	/transl_table=11		

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

Dictyoglomus thermophilum.
Dictyoglomus thermophilum
Bacteria; Dictyoglomus group; Dictyoglomus.
1 (bases 1 to 1423)
Reeves,R.A., Gibbs,M.D. and Bergquist,P.L..
Sequence of a beta-mannanase from the extreme thermophile
Dictyoglomus thermophilum Rt46B.1
Unpublished
2 (bases 1 to 1423)
Reeves,R.A., Gibbs,M.D. and Bergquist,P.L..
Direct Submission
Submitted (14-JUL-1997) Biological Sciences, Macquarie University,
Herring Road, Sydney, NSW 2109, Australia
Location/Qualifiers
1..1423
/organism="Dictyoglomus thermophilum"
/strain="Rt46B.1"
/db_xref="taxon:14"
101..1297
/gene="manaA"
101..1297
/gene="manaA"
/EC_number="3.2.1.78"
/note="ManA; mannan endo-1,4-beta-mannosidase; belongs to
glycosyl hydrolase family 26"
/codon_start=1
/transl_table=11
/product="beta-mannanase"
/protein_id="AA882454.1"
/db_xref="GI:2582053"
/translation="MHLEIIIGYAAPYKENSLYVNGEFTQNVKFPQSOKFTTVYAGL
IPUKNKTISIVKSGWFLIDFKIKAEIPTMNTKLVTNPSPKEAQKLDYLVS
IYKYTLISGOMGYKDAFMTWNITDKPEACFGCDMDYSRSRVGGASSRDVAIDDW
NMGGIVQFOHHNPNPKGLDTPCKWRMGFYTNATSFDEIYNHPSSDEVDKLIIRDI
DALAVQLRLQEALPILMRPLHEAEGRFWMGAKEACKKLRLFLFDRLVNYHKIN
NLAWVVTTSDPALKWYPCDEXYDIVGADYILKDXYDSPSTGMFNIVKLFGGKKLV
ALTENGIIPDDLMEKQAYWFMWTWSGFENDPNKNEISHIKKVNFHVITKDEL
NLKVES"

BASE COUNT	497	a	224	c	282	g	420	t
ORIGIN								

Query Match 8.1%; Score 327; DB 1; Length 1423;
Best Local Similarity 57.2%; Pred. No. 5.le-81;
Matches 682; Conservative 0; Mismatches 490; Indels 21; Gaps 4;

```

QY 2091 gcaaggctttacaattcacaatcggtacgcgttcctctcatgatgacaacgcagaa 2150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   94 GTCTGGAATCAGCAATTAATAATTGGATATGTCTGCACCCTATGGATATAAGGAAAATTC 153

QY 2151 ttctcttaaatggcaaacgctggagagctggacttgtaaagcggctgatttaa 2210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   154 CCTTATGTAANTGGAGAATTTCAAACCAATGTCAAATTTCCCAACTCTCAAAAATTTAC 213

QY 2211 agaaacgtccgggtggaagaagtcttgtgaatgcaggggcggaatacacatcggtttgaaac 2270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   214 AACCGTATATGCTGGTTTAATTCCTTTAAAAAATGGAAAAAATACAATAGTATAGTAAA 273

QY 2271 agctcgggcttggtatgatatacagctacgtcgagccagctgctgaccgtccac 2330
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   274 AACCTGGGGATGGTTTCTCTTGACTACTTTAAATCAAAAAGGCCAGAAATTCCTACCAT 333

QY 2331 gcatacggtaacgaaaacgctcataatccgaatgcagcagtagaagaacaaacattgat 2390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   334 GAATCCT--ACAACAATTAGTACACCCCTAATTCATCAAAAGAGGCCCAAAAATAAT 390

QY 2391 gaactacctgggtgatcaatacaggggaagaatatgctttccggtccaagagatatgccga 2450
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   391 GGACTATTAGTCAGTATATATGAAAGTATACTCTCTCGGGTCACATGGGATATAAGA 450

QY 2451 aattgattggcttcagcgcaatgtaggtaaaaagccgcttatcgcggcacttgacctgat 2510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

96291679

MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

2 (bases 1 to 3900)
Gibbs,W.D.
Direct Submission
Submitted (31-OCT-1995) Moreland D. Gibbs, School of Bio Sci,
Macquarie University, Sydney, NSW 2109, Australia
Location/Qualifiers
1..3900
/organism="Caldicellulosiruptor saccharolyticus"
/strain="Rt88.4"
/db_xref="taxon:44001"
/note="thermophilic anaerobe"
819..3554
/gene="manA"
819..3554
/gene="manA"
/EC_number="3.2.1.78"
/codon_start=1
/transl_table=11
/product="beta-1,4-mannanase"
/protein_id="AAC44232.1"
/db_xref="GI:1491795"

gene

CDS

DFDGTWMSGEAMGDSLKCTIKYVSQDLQRPNGKVALRLDVEFPNNGWQDGLGT
WFGVWEGQDFTCYKSVFEMFTPYDEFKSKSGQFPYKVIVINDKELGSEFNITAN
AGKVKINGDYTVIHKAFALPEDFTKRAQLVFOFAGONSXKGPILYLDNVRIRPE
DASNLSDGEQAEQVAVEDFTGVLVYFPGQKSFVNFIDIMFGFKYSQDGFN
KKTSLSTSQLKRSNGNGLVNASLAGTADFNIAIKLDRDKDKDFSKYSTLE
YTLVNPDKISGLMVASIDNPWQIIDFTALYNKDKNAIOIKINGKOYAVIKCSDN
LYNVTNKLRLIAGSYVYKTYPIIDNVKLIVAGKVPKVTPTSSIPNKNYRV
KIEATKSDVWAGYSEKENAKYSGKGYILLFGNNMNTLYDLKPKTYVIFISST
LGMVNSGDVWIDGELKGAQVNVKGFQEVVKKIYLTQGEHTLSQKSGYTI
AIDFTIEELVLANKNISVDTKLYTPNPNNAQKLMKYLASYGKILSGQSSDGG
KEIMIDVTPAVRGDFDPMDSRPSRGHTGTDEEAIKRWKSGGLVAFCHWN
APTLDOPKRWGRGYTEATDIDKAMDNPSTYKILRIDIAIABQLRLQKE
GVPVLFRLPESAGQFWGAKGPEYIKLWKLDFNLVNLNLIWVWNGQDAW
YPGNYVDLIDIEYEKAQYSPVADRFALKVYNARKMIALTECTGIPDPAVLKOE
GISWSFSVWAGNVMTGSKYNDENNDHMLKIYNSDYVITRDELDPKINPLK"

BASE COUNT
ORIGIN

1346 a 630 c 864 g 1060 t

Query Match
Best Local Similarity
Matches 506; Conservative 0; Mismatches 375; Indels 10; Gaps 2;

QY 2344 aaacgctcaatccgaatcgacagtagaagcaaaagcattgatgaactactgtgt 2403
DB 2580 ACAAGCTGTGAACCAACCAATCCATCTAATGCGCAAAAGCTTATGAATACCTGGCA 2639
QY 2404 gatcaatcagggaagaatattgtttccgggtcaagag-----gatatgcccgaatt 2454
DB 2640 AGCATATATGTTGAAGAATCTCTCTGGCAACAGACGAGTGGTGTGAGAAAGAAAT 2699
QY 2455 gattgcttcaagcaatgtagtataaagccgctattgpgcacttgacctgaattgac 2514
DB 2700 CAAATGATATTTGATTGACAAAAGATATCTCTGTCAGGGGCTTTGACTTTATGAC 2759
QY 2515 tattcccaagcagacggaacacggtcttagttccacagacgaggaagagcgattgaa 2574
DB 2760 TACTCACCAGCAGGCTGAGCATGGCACAAAAGGTACTGATGTAGAGAGGCAATAAA 2819
QY 2575 tgggataaagcaggggattgttaccctttgcatggcactggaacgcccgaaggtctg 2634
DB 2820 TGGTGAAGAGGGGTGGAATTTGCTTCTGCTGGCATGGAATTCACCAACAGGGCTT 2879
QY 2635 atcgatacagcaggaagaaatggtggagagcttctatgcatgattcattcattcatt 2694
DB 2880 ATTGACCCGCGGGAAGAAATGGTGGAGAGGATTTTACAGACAGCAACACATTTGAT 2939
QY 2695 atagatatgcatgaatcatcagagtcagagattataattgttatttcgacatc 2754
DB 2940 ATAAAGAGCTATGGACAACTCTAAATTCACCTGAGTATAGCTCATTTTGNAGATATA 2999

QY 2755 gatgtattgacgggcaattgaagaagttgcaggatgcgaagttcctgtctcgttccgt 2814
DB 3000 GACGCAATTCCTGAGCAGCTCAAGAGGCTTCAGAAAGAGGAGTCCAGCTCTTTTCAGA 3059
QY 2815 ccttgcacgaacgagggcaaatggttctgtggtggtggtggtggtggtggtggtggtggt 2874
DB 3060 CCCTGCACGAAGCTTCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3119
QY 2875 aaaaaattgtattttgatgcacgactggtttgactaatgtgcacaaattggaacattctg 2934
DB 3120 ATCAAACTCTGGAACCTTATGTTGATAGGCTTGTAAATATCACAAGTTAAACAACCTC 3179
QY 2935 atctgggtctggaactctgttctccgactggtatccggagagatagtatgtggtatt 2994
DB 3180 ATCTGGGTATGGAACGGTCAGGATGAGCATGTGTATCCGGGGGCAACATATGTGTGATA 3239
QY 2995 ttgagcttcgactcttccgacgagcgactacagccgcaaatgcaaaatgaa 3054
DB 3240 ATTGGCGAAGACATTTATGAGGAAAAAGCTCAATATATCCCTTATGCAGACAGTTTGA 3299
QY 3055 gacctgttaccttgggcaaggacaaaagctagttgtccatgagcgaacacgacctat 3114
DB 3300 -AAGGCATTCGAAGTACACAAATGCAAGAAAGATGATAGCACTTTACTGCTGGGAAAT 3358
QY 3115 cccggaccgactctgtaggagcgctatcaagcccttgagctggttcgctacatggtta 3174
DB 3359 TCCAGACCCAGCAGTATTAAAAAGGAGGATTTTATGAGGCTGGTTCCTCTCTGCGGC 3418
QY 3175 tggggtattcttgagacgagcgaacaaacacagctccttgagcatttga 3225
DB 3419 AGGAACTTTGTCATGACAGCGACGACGAGTATATGATGAGTGAATGACAA 3469

RESULT 13

RFU08621
LOCUS 2492 bp DNA linear BCT 04-APR-1996
DEFINITION Ruminococcus flavefaciens FD-1 CELB (celB) gene, complete cds.
ACCESSION U08621
VERSION U08621.1 GI:736356
KEYWORDS
SOURCE Ruminococcus flavefaciens.
ORGANISM Ruminococcus flavefaciens
Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Ruminococcus.

REFERENCE 1 (bases 1 to 2492)
AUTHORS Vercoe, P.E., Finks, J.L. and White, B.A.
TITLE DNA sequence and transcriptional characterization of a
beta-glucanase gene (celB) from Ruminococcus flavefaciens FD-1
Can. J. Microbiol. 41 (10), 869-876 (1995)
JOURNAL 96025594
MEDLINE 2 (bases 1 to 2492)
REFERENCE White, B.A.
AUTHORS Direct Submission
TITLE Submitted (11-APR-1994) Bryan A. White, Department of Animal
Sciences, University of Illinois at Urbana-Champaign, 1207 West
Gregory Drive, Urbana, IL 61801, USA
COMMENT On Mar 29, 1995 this sequence version replaced gi:475751.
FEATURES
Location/Qualifiers
1..2492
/organism="Ruminococcus flavefaciens"
/strain="FD-1"
/db_xref="taxon:1265"
/clone="celB"
550..2448
/gene="celB"
550..2448
/gene="celB"
/function="endoglucanase"
/codon_start=1
/transl_table=11
/product="CELB"
/protein_id="AAA95959.1"

source

gene

CDS

/db_xref="GI:736357"
translation="MRKSTAPLALAMAGOTLSAPLGLVLPETKYNAGGPMNKKVDL
KGERKEISPLIYQNTTDLKSVKTTAVRGNNRTAYWENNAAGSDHSHSDN
NLSDNPPAEVYVQLRSEAAKYGVYKMTTLMAGYVSADKCTGVKEDVAPSRKNE
VKFTKGAPDEPDLTGVVYMDYVNYIINKLSDQSPTGIOGYSLDNEPVLNTH
PRVHPPEVTTELGNKSIELAKAVKLPDPAEIPGALYGTAFDLDDDEQHTESGD
VKSNNYHWDLCYLDOMKASBEERGLDVLDIHYSESARTGAEDRVQSVRTIYE
EGFSESWIGOMCQMONVPIIPTIKKSIDTYPPGTLKLAISEYKFGEDSGTITQAQEA
LGCFADQGVYLATLWGEPEFIISGLNLYNDKGGCFGDTLIPASTEDVSKSTVAA
VNDGDESKVMTINKNNMNAENAVIDLENAKDYKSAAYVAVYGNDOVRLDIDVKD
VKDNKVNVELPASAAMVYVSDDAAFDGEKIVYEKKVTEKFEEDKPSMINKNYV
EIPITDPEQYVKVINGDVTSSAGSGWATRLCSMHQKQGGQGGFIDIQLOSPAGQ
GSVRYRVK"

BASE COUNT 725 a 572 c 621 g 574 t
ORIGIN

Query Match 4.8%; Score 193.6; DB 1; Length 2492;
Best Local Similarity 50.6%; Pred. No. 3.6e-43;
Matches 706; Conservative 0; Mismatches 94; Indels 96; Gaps 6;

Qy 132 tttaccattaaagtcgatacatcaaggatcgtaagcctattagccttattattacgg 191
Db 660 TATGAATATCAAGGTAGACCTCAAGGGGAGCGCAAGAGATAGCCCTCTTATCAGG 719

Qy 192 tacgaatcaggagtgccagcgatgag-----aatctgactgccagacgacttggtg 245
Db 720 CGTTAACCAGTATATCCACAGACCTCAAGAGCGTCAAGACTACTGCTAGCTCAGGCGG 779

Qy 246 caatcgatcagcgatataatcgggaaacaataatcgaatcgaagcagcagcagcagc 305
Db 780 TAACCGTATGACTGCTTACACTCGGAGACATGCTTCAACGACGAGTTCGGATGGA 839

Qy 306 gcagtcacgagatagctatttatcgacaacgcgcggattgacaaaagcgaatgtgaaa 365
Db 840 ACACAGTTCGATTAACAACCTCT-----CGGATTCAAAATCC 875

Qy 366 gccaggtgcggtgcaacctggttcacgatcaatcgtaagcagcgagcagcattcttt 425
Db 876 TCCTGCTGAGTGTTCAGAGACTTTCAGAGGAGCTGCAAGAGTACGGCTTGATTACA 935

Qy 426 agtcaca---ctgcgcatgcccgtgttatgtggccaaagatggaagcgaagtgccagga 482
Db 936 AATGACAACATCTCCAGATGCGAGCTAGCTTTCAGCTGATGAAGGACGCGACTGTCAGGA 995

Qy 483 aagcgaagagcctctccgctcgtggtggaatgaggtgtaaacgcgtataaaatgcgcgtt 542
Db 996 AGATGAAGTCTGCTCTTCAAGGCGCTGGAACGAGGTAAAGTTCAAAAGGGCGCACCTTT 1055

Qy 543 tcaattgcagcctgactcgaagacaatcaggttttatcgatggaattgcgaacttttt 602
Db 1056 TGCTGATGAGCCGTGATCTGACAGCGCGTGTGTTATATGGATGAGTACGTAAACTACAT 1115

Qy 603 agtgaaaaagtcagcggtgtgcttcaaaaaacggcggtgaaaggtactcgcctgcacaa 662
Db 1116 CATCAATAAGCTGGCGGATTCACAGTACACCTACAGGTATCCAGGGCTACAGCTTGATTA 1175

Qy 663 tgaaccgctctctgctgcatacgcgcgcattcattggtgaaaggtcgagcgaa 722
Db 1176 CGAGCCTGTCTCTGGAAACGACTACTCACCAGAGGTATCACCAGGCTGTAACCTATCGA 1235

Qy 723 agatgtggtagacggtgcggtlaagtttatccaaagcgcgttaagcggttgacgcggtgc 782
Db 1236 AGAGCTTGGAAACAAGTCCATCGAATCTGTAAGGCTGTCAAGAGCTTGATCCTAAGGC 1295

Qy 783 ggaatttttggcccgctt-----ttacgggttttggccctataaaga 827
Db 1296 TGAGATCTTCGGACCTGCACTCTATGGCTACACTGCTTTCGATCACCTCGACGATGACA 1355

Qy 828 tcitcaaacatgcacgtgattggaactctgtataaagcgaactacagctggttctggaacta 887
Db 1356 GCAGCATACAGATCGGGAGACGCTAAAGTCCAGAACAACTATCACTGTGATCTGAGCTG 1415

Qy 888 ttacctcgatcaaatgcgcctcagctcgcaagccggaagccgaagagattgctgctgattctt 947
Db 1416 CTATCTCGACCAGATGAAGAGGCTTCTGAAGAGAGGACCCCGTCTCTCTCGATGACT 1475

Qy 948 cgtgtacacgtgtactcctgaagcgtatggcgaggagcagacatacgaattacaagttaggtag 1007
Db 1476 TGACATCCACTACTACTCGGAATCTCAAGAACAGC----- 1512

Qy 1008 caacgacgaagaagaagcgaagaatgcaagcgcctcgtactcttgggagcgcaccta 1067
Db 1513 -----GCTGAGGACAGAGATTCACTCAGTTCGTCTCTCTATGAGGAAGGCTT 1559

Qy 1068 caaggaaagtagctgagctcgctcaatggaaacagtggaattcttgccttactgcctcgatt 1127
Db 1560 CTCCGAAACAGCTGGTCACTGGTGTATGACAGATGTCCTATCTTCCGACTAT 1619

Qy 1128 aaagcagtcggtgggataagattattaccgggggaaacaaagctggtttgactgagtagta 1187
Db 1620 AAAGAAGTCCATAGATACATACTATCCGGCACAAAGCTGGCTATCTCAGAGTATAACTT 1679

Qy 1188 tgggtgcgaaaaatgattattccgcggttatcgtatggtcgcgagtggtggtggtggttggg 1247
Db 1680 CAAGGCGGGAGGAGTACTTCCGGTACTATCGCTCAGGAGGAGGACTGGGATGCTTCCG 1739

Qy 1248 caaaaaacgagctttatttatggcaaaactactggaagtttaaaggatgggtgccacaactaagt 1307
Db 1740 TGATCAGGGCGTATATCTCGAACCTCTCTGG-----GGGGTGACCCATTCTAT 1787

Qy 1308 tagtgcgctttacagctttacgcgaattatgacgggaaagcttacttacttctggtgatat 1367
Db 1788 CATCTCAGTATCAACCTCTACACTAACTACGACGCGGAGGCGGATGCTTCCGTGATAC 1847

Qy 1368 cagcgttctatgcgcaaacgctcggtattgttaagtcggtgcgctgcttccgtaacgga 1427
Db 1848 TCTCATCCCTGCTTCTACTGGAAGATGTATCAAGTCCAGCACATAGCTGCTGTTAACA 1907

Qy 1428 tgcactcctcaagaactgcacctcgttgcataagataaaagcagtcgagcagtcattcga 1487
Db 1908 CGGCGATGATGACAAAGGTAACTGTATGATGATCACCACCAAGATATACAGAAAGCTGAGA 1967

Qy 1488 cgcccaatttgattt 1503
Db 1968 CGCTGTTATCGACCTT 1983

RESULT 14
RMRMANAG Locus RMRMANAG 4831 bp DNA linear BCT 16-JUN-2000

DEFINITION R.marinus manA gene.
ACCESSION X90947
VERSION X90947.1 GI:2832599
KEYWORDS endo-1,4-beta-mannanase; manA gene.
SOURCE Rhodothermus marinus.
ORGANISM Rhodothermus marinus.
REFERENCE 1 (bases 1 to 4831)
AUTHORS Politz,O., Krah,M., Thomsen,K. and Borris,R.
TITLE A highly thermostable endo-(1,4)-beta-mannanase from the marine bacterium Rhodothermus marinus
JOURNAL Appl. Microbiol. Biotechnol. 53 (6), 715-721 (2000)
MEDLINE 20373941
REFERENCE 2 (bases 1 to 4831)
AUTHORS Politz,O.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1995) O. Politz, Humboldt University of Berlin, Institute of Biology, Warschauer Str. 43, D-10243 Berlin, FRG
REMARK 3 (bases 1 to 4831)
AUTHORS Politz,O.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1998) O. Politz, Dermatology, Benjamin Franklin Medical Center The Free University of Berlin, Hindenburgdamm 30, Berlin 12200, Germany

Tue Jul 16 15:19:48 2002

us-09-784-554b-3.rge

Page 19

Search completed: July 12, 2002, 18:46:32
Job time: 32779 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 19:10:52 ; Search time 1513.98 Seconds
(without alignments)
4599.668 Million cell updates/sec

Title: US-09-784-554B-3

Perfect score: 4056

Sequence: 1 atgaagcgaaataatagtag.....ccgaaccgacagcagtaattaa 4056

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802:*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4055.2	100.0	4056	22	AAH75060
2	3374	83.2	4059	22	AAH75059
3	1721.4	42.4	2141	22	AAH75061
4	630.2	15.5	2977	12	AAQ13001
5	629.4	15.5	1854	12	AAQ14842
6	417.4	10.3	6415	20	AAH55662
7	417.4	10.3	6416	24	AAH26526
8	399.2	9.8	2445	21	AAZ45347
9	378.4	9.3	1488	21	AAZ45348

10	327.2	8.1	1464	21	AA245341	DNA encoding a Hum
11	247.2	6.1	3041	21	AA245350	DNA encoding a man
12	133.2	3.3	2600	12	AAQ15178	Portion of pAEC-1
13	90	2.2	11706	20	AAH55661	DNA sequence encod
14	87	2.2	11707	24	AAH26525	Active cellulase p
15	85.2	2.1	2175	14	AAQ49820	NK-1 cellulase gen
16	85.2	2.1	2029	20	AAH55660	DNA sequence encod
17	85.2	2.1	2029	24	AAH26568	Active cellulase h
18	61.4	1.5	1775	12	AAQ13003	Endo3 gene encodin
19	54	1.3	2892	18	AAH23276	Cyclic-isomaltoligo
20	51.2	1.3	1461	10	AAH91621	Beta-mannase gene
21	46	1.1	4839	11	AAQ06827	Alpha amylase pull
22	45.6	1.1	1011	21	AAZ29313	B. subtilis strain
23	45.6	1.1	1029	21	AAZ29847	Bacillus subtilisi
24	44.2	1.1	5059	20	AAH84332	Stealth virus nucl
25	42.8	1.1	1086	21	AAZ45349	DNA encoding a Bac
26	41	1.0	6741	21	AAH10595	Gene encoding a su
27	40.2	1.0	14041	22	AAH48024	Internal control B
28	39.6	1.0	4749	23	ABL14248	Drosophila melanog
29	39.4	1.0	10732	21	AAH10594	Gene encoding a su
30	38.4	0.9	4590	22	AAH24065	Yeast AOD9604-asso
31	38.2	0.9	1812	23	ABL14249	Drosophila melanog
32	37.6	0.9	6192	23	ABL10254	Drosophila melanog
33	37.4	0.9	2916	17	AAH15957	Cycloisomaltoligo
34	37.4	0.9	2916	20	AAH10902	Bacillus citase co
35	37	0.9	243	22	ABA49895	Human breast cell
36	37	0.9	243	22	ABA67813	Human foetal liver
37	37	0.9	243	22	ABA34872	Probe #13338 for g
38	37	0.9	243	22	AAH41969	Human bone marrow
39	37	0.9	243	22	AAI22735	Probe #12668 for g
40	37	0.9	243	22	AAI48035	Probe #16721 used
41	37	0.9	243	22	AAI08403	Probe #8394 used t
42	37	0.9	477	22	ABA44753	Human breast cell
43	37	0.9	477	22	ABA55208	Human foetal liver
44	37	0.9	477	22	ABA24955	Probe #3421 for ge
45	37	0.9	477	22	AAK28923	Human bone marrow

ALIGNMENTS

RESULT	1
AAH75060	DNA encoding a man
ID	AAH75060 standard; DNA; 4056 BP.
XX	AAH75060;
AC	AAH75060;
XX	29-OCT-2001 (first entry)
DT	Nucleotide sequence of xyloglucanase enzyme.
XX	xyloglucanase; family 44; glycosyl hydrolase; detergent;
XX	cellulosic fiber; textile scouring; ss.
OS	Paenibacillus polymyxa.
XX	Key
XX	Location/Qualifiers
FT	CDS
FT	1..4056
FT	/*tag= a
FT	/product= "xyloglucanase"
FT	/transl_except= "(pos: 4039..4044, aa: Glu)"
XX	WO200162903-A1.
PN	30-AUG-2001.
XX	21-FEB-2001; 2001WO-DK00116.
XX	24-FEB-2000; 2000DK-0000291.
XX	(NOVO) NOVOZYMES AS.
XX	Schnorr K, Jorgensen PL, Schuelein M;

XX WPI: 2001-522819/57.
 DR P-PSDB: AAG63963.
 XX
 PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries
 XX
 PS Claim 14; Page 85-86; 97pp; English.
 XX
 CC The present sequence encodes a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
 CC linen.
 XX
 SQ Sequence 4056 BP; 1161 A; 888 C; 1070 G; 935 T; 2 other;

Query Match 100.0%; Score 4055.2; DB 22; Length 4056;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4056; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgaagcgaaataatagtagtattgttccaaacgttccaaatggtcgtcgtc 60
 DB 1 atgaagcgaaataatagtagtattgttccaaacgttccaaatggtcgtcgtc 60
 QY 61 atgcatgcacgattatagtaggggtgctctaccgactccaaactgtagtccaggtcaa 120
 DB 61 atggcagcagataagtaggggggctctaccgactccaaactgtagtccaggtcaa 120
 QY 121 acggcaaaagactgttaccattaaagtcgtacatccaaagcgtgtaagcctattagccct 180
 DB 121 acggcaaaagactgttaccattaaagtcgtacatccaaagcgtgtaagcctattagccct 180
 QY 181 tatattacgtacgaatcaggaatggtgagggcgatgagaatctgactgccaagacactt 240
 DB 181 tatattacgtacgaatcaggaatggtgagggcgatgagaatctgactgccaagacactt 240
 QY 241 ggtggcaatcgaatgacggatataactgggaaacaaatattgtccaatgcaggaagcgac 300
 DB 241 ggtggcaatcgaatgacggatataactgggaaacaaatattgtccaatgcaggaagcgac 300
 QY 301 tggatgcagtcacgtagctatttatgcgcaaacgcggattgacaaagccgaatgt 360
 DB 301 tggatgcagtcacgtagctatttatgcgcaaacgcggattgacaaagccgaatgt 360
 QY 361 gaaagccaggtgcggtggcaacctgtttcagatcaaatcgtgagcaggacacatat 420
 DB 361 gaaagccaggtgcggtggcaacctgtttcagatcaaatcgtgagcaggacacatat 420
 QY 421 tcttagtcacactgcgtagtggcgggttatgtggcgaagcgtggaacggaagtggtcag 480
 DB 421 tcttagtcacactgcgtagtggcgggttatgtggcgaagcgtggaacggaagtggtcag 480
 QY 481 gaaagcgaaggtccttcgtcgtggtggaatgaggtcgttaaacgtcaaaatgcgcg 540
 DB 481 gaaagcgaaggtccttcgtcgtggtggaatgaggtcgttaaacgtcaaaatgcgcg 540
 QY 541 ttccaattgcagctgactgaagacaatcaggtttatgcggatgaattcgaacttt 600
 DB 541 ttccaattgcagctgactgaagacaatcaggtttatgcggatgaattcgaacttt 600
 QY 601 tttagtgaagaaaglacggcgtgtgttccaaacaaacggcggtgaaaggtactcgtcgcac 660
 DB 601 tttagtgaagaaaglacggcgtgtgttccaaacaaacggcggtgaaaggtactcgtcgcac 660

DB 601 tttagtgaagaaaglacggcgtgtgttccaaacaaacggcggtgaaaggtactcgtcgcac 660
 QY 661 aatgaaccgctctctggtcgcatacgcattccgcgcattcattcgtgtaaaaggtcggagcg 720
 DB 661 aatgaaccgctctctggtcgcatacgcattccgcgcattcattcgtgtaaaaggtcggagcg 720
 QY 721 aaagagtgtgtagaccggtcggttaagtattatccaaagccgctaaagcggttgaacgcggt 780
 DB 721 aaagagtgtgtagaccggtcggttaagtattatccaaagccgctaaagcggttgaacgcggt 780
 QY 781 ggggaatttttgggcccgttttaccggttttggcgccataaaagatctcaaaactgca 840
 DB 781 ggggaatttttgggcccgttttaccggttttggcgccataaaagatctcaaaactgca 840
 QY 841 cctgattggaactctgtaaaagcgaactacagctggttcgtggaactattacactcgaatca 900
 DB 841 cctgattggaactctgtaaaagcgaactacagctggttcgtggaactattacactcgaatca 900
 QY 901 atgctcctcagctcgcaaacggaagcgaagattgctggtgatttcttcgattacactgg 960
 DB 901 atgctcctcagctcgcaaacggaagcgaagattgctggtgatttcttcgattacactgg 960
 QY 961 tatcctgaagcagatggcgaggacatacgaattacaaatgaggtaggcaaacgacaaaacg 1020
 DB 961 tatcctgaagcagatggcgaggacatacgaattacaaatgaggtaggcaaacgacaaaacg 1020
 QY 1021 aagaaagccagaaatcgaagcgcctcgttcttgggtggtacccgacccataaagaaagatagc 1080
 DB 1021 aagaaagccagaaatcgaagcgcctcgttcttgggtggtacccgacccataaagaaagatagc 1080
 QY 1081 tggatgcctcaatggaaacagtgaaattcttgccttactgcctcgtatggttcttcgattc 1140
 DB 1081 tggatgcctcaatggaaacagtgaaattcttgccttactgcctcgtatggttcttcgattc 1140
 QY 1141 gataagtatccgggggaacaaagcgtggttgcagtagtagtatactggtggcgaaaaat 1200
 DB 1141 gataagtatccgggggaacaaagcgtggttgcagtagtagtatactggtggcgaaaaat 1200
 QY 1201 gatattccggcggtgctcgtatgctggtggtggtggtggtggtggtggtggtggtggtggt 1260
 DB 1201 gatattccggcggtgctcgtatgctggtggtggtggtggtggtggtggtggtggtggt 1260
 QY 1261 tataatggcaaaactgctggaagttaaaggtggtggtggtggtggtggtggtggtggtggtggt 1320
 DB 1261 tataatggcaaaactgctggaagttaaaggtggtggtggtggtggtggtggtggtggtggtggt 1320
 QY 1321 aagctttaccgcaaatatgacgggaaagcgtccttcttcgtggtggtggtggtggtggtggtggt 1380
 DB 1321 aagctttaccgcaaatatgacgggaaagcgtccttcttcgtggtggtggtggtggtggtggtggt 1380
 QY 1381 caaacgtcggtattgttaatagctcgtggtggtggtggtggtggtggtggtggtggtggtggt 1440
 DB 1381 caaacgtcggtattgttaatagctcgtggtggtggtggtggtggtggtggtggtggtggtggt 1440
 QY 1441 gaactgcacctcgttgcataaataaagcgtggtggtggtggtggtggtggtggtggtggtggtggt 1500
 DB 1441 gaactgcacctcgttgcataaataaagcgtggtggtggtggtggtggtggtggtggtggtggtggt 1500
 QY 1501 ctttcggcgagacacttaccggttccggtggtggtggtggtggtggtggtggtggtggtggtggt 1560
 DB 1501 ctttcggcgagacacttaccggttccggtggtggtggtggtggtggtggtggtggtggtggtggt 1560
 QY 1561 caaattaaggagcagcgcccaatcscaaaatttcaggcaacccggttttaccctatacagta 1620
 DB 1561 caaattaaggagcagcgcccaatcscaaaatttcaggcaacccggttttaccctatacagta 1620
 QY 1621 cgcctttgacgggttatacacaatcgttgcagcgggaatgatacaccctgtgagaaaaat 1680
 DB 1621 cgcctttgacgggttatacacaatcgttgcagcgggaatgatacaccctgtgagaaaaat 1680
 QY 1681 cctgaaagcttgcgtgaggtgagcgtggtggtggtggtggtggtggtggtggtggtggtggtggt 1740
 DB 1681 cctgaaagcttgcgtgaggtgagcgtggtggtggtggtggtggtggtggtggtggtggtggtggt 1740

|||||
Db 3901 attcaactcgattataaagacagactggtcgactataacgaagtgagattactcg 3960
QY 3961 tataaagtgacagacttctacacgatacgaagtgactgtatcacaaatggc 4020
Db 3961 tataaagtgacagacttctacacgatacgaagtgactgtatcacaaatggc 4020
QY 4021 gtactgtttggggaacgaacgacagactaattaa 4056
Db 4021 gtactgtttggggaacgaacgacagactaattaa 4056
RESULT 2
AAH75059
ID AAH75059 standard; DNA; 4059 BP.
XX
AC AAH75059;
XX
DT 29-OCT-2001 (first entry)
DE Nucleotide sequence of xyloglucanase enzyme.
DE
KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
KW cellulosic fiber; textile scouring; ss.
XX
OS Paenibacillus polymyxa.
XX
FH Key Location/Qualifiers
FT CDS 1..4059
FT /*tag= a
FT /product= "xyloglucanase"
XX
PN WO200162903-A1.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-DR000116.
XX
PR 24-FEB-2000; 2000DR-0000291.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Schnorr K, Jorgensen PL, Schuelein M;
XX
DR WPI; 2001-522819/57.
DR P-PSDB; AAG63962.
XX
PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
PT useful for detergent compositions, and textile or cellulose fiber
PT processing industries
XX
PS Claim 14; Page 80-81; 97pp; English.
XX
CC The present sequence encodes a xyloglucanase of the invention. The
CC specification describes a xyloglucanase enzyme belonging to family 44
CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
CC detergent compositions and prevents binding of certain soils to the
CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
CC room temperature and has a half life of more than 50 days when incubated
CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
CC enzyme is used in detergent compositions, textile industry for improving
CC the properties of cellulosic fibers, yarn, and woven or non-woven
CC fabrics, preferably in textile scouring process, and in cellulose fiber
CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
XX
SQ Sequence 4059 BP; 1187 A; 893 C; 1046 G; 933 T; 0 other;
Query Match 83.2%; Score 3374; DB 22; Length 4059;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 3664; Conservative 2; Mismatches 387; Indels 9; Gaps 5;

QY 1 atgaaggcgaaaaatagtagtagtatttggctccaaacggttccaaatggctgctgctc 60
Db 1 atgaaggcgaaaaatagtagtagtatttggctccaaacggttccaaatggctgctgctc 60
QY 61 atggatgcacgattattatagtagtaggggtgctctaccgactccaaactgtatgtcacggtcaa 120
Db 61 atggctgcacgattagtagtaggggggttaccctgctccagctggtgtcacggtcaa 120
QY 121 acggcaagactgtaccattaaagtcgatacattcaagagatcgtaagcctattagccct 180
Db 121 acggcaagactgtaccattaaagtcgatacattcaagagatcgtaagcctattagccct 180
QY 181 tatattacgggtacgaatcaggaggttggcaggcgatgagaaatctgtactgtccagagactt 240
Db 181 tatatacgggtacgaatcaggaggttggcaggcgatgagaaatctgtactgtccagagactt 240
QY 241 ggtggcaatcgaatgacccggatataaactgggaaacaatattgtccaatgcaggaagcgac 300
Db 241 ggtggcaatcgaatgacccggatataaactgggaaacaatattgtccaatgcaggaagcgac 300
QY 301 tggatgcagtcacgagcagtagctatttgcgacacgcgcgattgacaaaagcgaatgt 360
Db 301 tggcagcaatctagcgataactatttgcagtaattgtggtgactgacacaaagcgaatgt 360
QY 361 gaaaagccagggtgcgggtggcaacctctgttcacgactcgaatcgaatgcagggcacat 420
Db 361 gaaaagccagggtgcgggtgcagactctgttcacgactcgaatcgaatgcagggcacat 420
QY 421 tcttagtcacactgcgagtcgggttatgtgccaagagatggaaacggaagtgacag 480
Db 421 tcttagttacgttccgagtcgggttatgtgccaagagatggaaacggaagtgacag 480
QY 481 gaaagcgaaaaggtctctccgctcgggtggaatgaggtcgtaaacgcgttaaaatgcgcg 540
Db 481 gaaagcgaaaaggtctctccgctcgggtggaatgaggtcgtaaacgcgttaaaatgcgcg 540
QY 541 ttccaattgcagcctgatctgaaagacacatcaggtttatgcggatgaaattcgtcaacttt 600
Db 541 ttccaactacagcctgatctgaaatgacacatcgggtctatgtggaatgagttcgtccatttt 600
QY 601 ttagtgaanaagtagcgggttcttcaacaaacggcggtgaaaggtacgactcgtcgac 660
Db 601 ttagtgaanaagtagcgggttcttcaacaaacggcggtgaaaggtacgactcgtcgac 660
QY 661 aatgaacccgctctctggtcgacacgcacccacgacattcatggtgaaaggtcgagcg 720
Db 661 aatgaacccgctctctggtcgacacgcacccacgacattcatggtgaaaggtcgagcg 720
QY 721 aaagagttggtagacccggtcggttaagttatccaaagcgcgttaagggcgttgacgcgggt 780
Db 721 aaagagttggtagacccggtcggttaagttatccaaagcgcgttgaaagcgttgacgcgggt 780
QY 781 gcgggaattttggcccgcttcttaccggttttggcgctataaagatcttcaactgca 840
Db 781 gcagaggtttttggcccggttcttaccggttttggcgctataaagatcttcaactgca 840
QY 841 cctgattggaactctgtaaaagggcaactacacgctggttcgtggactattacctcgatcaa 900
Db 841 cctgattggaactctgtaaaagggcaactacacgctggttcgtggactattacctcgatcaa 900
QY 901 atgcgctcagctgcgaagcgaaggaagatgctggtgagtgcttcgtacatggtg 960
Db 901 atgcgctcagctgcgaagcgaaggaagatgctggtgagtgcttcgtacatggtg 960
QY 961 tatcctgaagcgatggcgaggcgatcgaatataatgaggtgagtagcagcgaagc 1020
Db 961 tatcctgaagcgatggcgaggcgatcgaatataatgaggtgagtagcagcgaagc 1020
QY 1021 aagaagccagaatgcgaagcgcctctgactttgtgggactccgacctacaaggaagatgc 1080
Db 1021 aagaagccagaatgcgaagcgcctctgactttgtgggactccgacctacaaggaagatgc 1080

Qy	1081	tgatcgctcaatgaacagtgaaattcttgcctttactgctcgattaaagcagtcggtg	1140
Db	1081	tgatcgctcaatgaacagcgagttttgcccatactactcgattgaagcagtcggtg	1140
Qy	1141	gataagtattaccgggaaccaagctggcttgactgagtatagctatggtggcgaatt	1200
Db	1141	gataagtattaccgggaaccaagctggcaatgacgagtatagctatggtggcgaatt	1200
Qy	1201	gatattccggcggtatcgctatggcgatgctggcgatcttggcgaacacgacgt	1260
Db	1201	gatattccggcggtatggcgatgacgagtgctgggtatcttggcgaacacgacgt	1260
Qy	1261	tatatggcaactactggaagttaaagtggtgccacaactacgttagtcgcgttac	1320
Db	1261	tatatggcaactactggaagttaaagtggtgccacaactacgttagtcgcgttac	1320
Qy	1321	aagctttaccgcaattatgacgggaacaaagctctactttcggtgatatcagcggtcaatgg	1380
Db	1321	aagctttaccgcaattatgacgggaacaaagctctactttcggtgatatcagcggttagtgcg	1380
Qy	1381	caaacgcggatattgttaatagctcggtgcatcttcgcgaacgagtgatcactacaaa	1440
Db	1381	caaacgcggatattgtcaatagctcggtccatgctctgtaacgaatgcatccgacaaa	1440
Qy	1441	gaactgcacctggttgcataataaagcatggacagtgcatctcgacgcccatttgat	1500
Db	1441	gaactgcacctggttgcataataaagcatggacagcgcatctcgacgcccatttgat	1500
Qy	1501	ctttccggcgagcagacttaacggttcocggtaaagtatggggtttcgacaaaaatagctcg	1560
Db	1501	ctttccggcgagcagacttaattccggtaaagtatggggtttcgacaaaaacagctcg	1560
Qy	1561	caaataaaggagcagcgccaatcacscaaatttcaggaacacggtttacctatacagta	1620
Db	1561	caaataaaggagcagcgccaatcaccaatttcaggaacacggtttacctatacagta	1620
Qy	1621	ccgcctttgacggcttatcacatcggttgactgcggcgcaatgata---cacctgtagaa	1677
Db	1621	ccgcctttgacggcatcacatgtgctgactctggcaatgacagcttcacgtggaac	1680
Qy	1678	aatcctgaaagctttgcctgagggctgaggtcggtcgat-ggaagtcgatttato--tg	1734
Db	1681	ggtcctgaaagctttaagctgaaagctgaggtggtgatgggaagtcctatttactctgg	1740
Qy	1735	gacgtcttcagcgaggtgtaggttacagctacagcgccacagctatgaaacggtcct	1794
Db	1741	gatgtcttcagcgaggtgtcaggaacacagctacagcgccacagctatgaaacggtcct	1800
Qy	1795	tttgctgctgtagcatccaaacttgctgaacactcttatacgggatacgaacgttaacgaac	1854
Db	1801	ttcacgtgtgtagcatccaaacttgacggaacgtcttatacgggatacgaacgtgacaaac	1860
Qy	1855	ggcaacttctattattataaaataaccgcaaaaacaaagacgggaacgacgcatccaat	1914
Db	1861	ggtactctactattacaagttaaccgcccacaaacaaataaggagatcgacggaatccaat	1920
Qy	1915	gtctgaaagcggttccgcggcgctgtagacggtccgagctgctatgaacggaagat	1974
Db	1921	attttgaaagcggttccggaagtgcctgtaaacgggtcccgctgctgtatgaacggaagaa	1980
Qy	1975	ggcagctgaaggggacggtgtggaatccagtggaacggcggtcttcctgactggttat	2034
Db	1981	ggcagctgaaggggaaccattgtggaatccagcgggacggcgtactcgggtgctggttat	2040
Qy	2035	gtaactaatttccaaatgcaagggattccctgcagatgacgatacaggtccacagcgga	2094
Db	2041	gtaacgaatttccaaatccagggattctctgcagatgacgatacaggtccacagcgga	2100
Qy	2095	ggctgtacaactctacaatcgataccggttctctctcatgatgacaacgcagaaatttc	2154
Db	2101	ggctgtacaactctacaatcgataccggttctctctcatgatgacaacgcagaaatttc	2160
Qy	2155	tcctttaaaggcaacgctctgagagctggttactttggaaaaacggctgattttaagaaga	2214

2161	Db	tcattaaacgcgaagcgtttggcgaaactgtcttaagaaaaacgcgtgattttaagaa	2220
2215	QY	acgtccggctggtaagcttctgttgatcaggggcgaaatcacgtctggtttgaaacaggc	2274
2221	Db	acttcggaggcaaggtgctgttgaaatcgagcgcgaatacgtatcagttttgaaacaggc	2280
2275	QY	tggggctggatgatcatcgactacgtcaagctggagccagctgctgaccgtccacgcgat	2334
2281	Db	tggggctggatcgatcgactacgtcaagctggagcctgcgctgaccgcccaactcat	2340
2335	QY	ggcgtaacgaaacgctcatcaatccgaatcgacagctagaagcaaaagcatgtgatgaac	2394
2341	Db	ggcgtaaccaaacgcttaccaatccgaatcgacgcgttagaagcaaaagcatgtgatgaac	2400
2395	QY	tacctggttgataatcacgggaagaatgctttccgctcaagaggatatgtcccgaaatt	2454
2401	Db	tactcgttgatcaatcacgggaagaatgctctctcgtcgaaggaggaaataaacgaaatt	2460
2455	QY	gattgcttcaagcgaatgagttaaaagcggctattgcgcgaactgacctgacctgattgac	2514
2461	Db	gattgcttcaagcgaatgagttaaaagcgcggattgcagcgtttgacctgacctgac	2520
2515	QY	tattcccaagcagagcggaacacaggtcttagttccacagagacggaagaaaggcgattgaa	2574
2521	Db	tattcccaagcagagcggaacacaggtcttagttccacagagcgagaagaaaggcgattgca	2580
2575	QY	tgggataagcaagggggcattgtacctttcatgacctggaacgcgcgcgaagagctcg	2634
2581	Db	tgggaataagcaagggggcattgtacctttcatggcactgggaacgcgcgaagagctcg	2640
2635	QY	atcgatcacgaggaagaaagaaatgggtgagaggtctctatgcgattcgcgactacattcgat	2694
2641	Db	atcgatcacgaggaagaaagaaatgggtgagaggtctctatgcgattcgcgactacattcgat	2700
2695	QY	atagaatatgcgatgaatcatccagagtcggaagattataaatgcttatcttcgcgcacatc	2754
2701	Db	atagaatatgcgatgaatcatccagagtcggaagattataaatgcttatcttcgcgcacatc	2760
2755	QY	gatgtattgcaggcgaattgaagaagtgcagagatcgaaagtctcctgtcctgttcctgt	2814
2761	Db	gatgtattgcagggcgaattgaagaagtgcagagatcgaaagtctcctgtcctgttcctgt	2820
2815	QY	cccttgcaagaaaggagggcgaatggttctgtgtgggcgtctaaaggctcctgagccctgtt	2874
2821	Db	cccttgcaagaaaggagaaatggttctgtgtgggcgtcccaaggctcctgagccctgtt	2880
2875	QY	aaaaattgtatatatttgatgcacgatcgttgactaatgtgcaacaaattgacaacatctg	2934
2881	Db	aaaaagctgatatatttaatgcacgacctgttgacgaatgtgcaacaaattgacaacatctg	2940
2935	QY	atctggctggaactctgtctccgactggtatccggagatcgagatggagtatgtggatatt	2994
2941	Db	atttgggtatggaaattctgtctcccgatgtggtatccggagacgagtatgtggatatt	3000
2995	QY	ttgagcttcgactcttatccgcaagcagcgactacagcccgaaattgcaaaatatgaa	3054
3001	Db	ttgagctttgactcttatccgcaagcaggtgattacagcccgcaaaatttcaaaatacgaa	3060
3055	QY	gaacttgttaccttgggcaaggaacaaaaagctagtttgccatgagcgaacacgacacctat	3114
3061	Db	gaacttgttaccttgggcaaggaacaaaaagctag-ttgccatgagcgaacacgacacctat	3119
3115	QY	cccgaccgcgactctgatgaagcgtatacaagcccatggagctggttcgctacatggta	3174
3120	Db	cccgacccttgaatttgatgaagggctatcaagctcatctggagctggttcgctacatggta	3179
3175	QY	tggggattctctgagagacggcaaacaaacacgctcctctggagcatttgaacaaagtgtat	3234
3180	Db	tggagatttttgagagacggcaaacaaacag--ccttgagcactcgaacaaagtgtat	3237
3235	QY	aatcatccgaacgtcattacgcttgaaaagctcccgactaaccttaaacgctatggcaatt	3294

Db 1398 cctgacgtgtgtaataaaatttcagcatccgcatcaacgcgtactttccagctgtctgg 1457
 QY 1509 cgaagcagcttacgttccggtataaagtatgggtttcgacaaaaataagctgcgaataaa 1568
 Db 1458 tgataaaacacacatccggagagatggtggttcgacaaacgcgagatccgacattac 1517
 QY 1569 ggaagcagcgcacacacacaaatttcagcaacgcgtttaccctatcacagcagcttt 1628
 Db 1518 ggaacagcagctataacgaattataacaaacaaatccacgtatcagctctctccatt 1577
 QY 1629 gacgcttatcacatcgtgtgtgactgcgcggcaatgatacacactgt 1673
 Db 1578 gtcggttaccacattgtctgaagcggatagcaccggaaccggt 1622

RESULT 6

AA55662 ID AA55662 standard; DNA; 6415 BP.

AC AA55662;

DT 30-JUL-1999 (first entry)

DE DNA sequence encoding truncated cellulases.

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 KW cotton-containing fabric; stonewashing; ss.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

PF 15-SEP-1998; 98EP-0810919.

XX 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK;

PI Gibbs MD, Morgan H, Williams DP;

XX WPI; 1999-315403/27.

DR P-PSDB; AAV13493.

PT New truncated cellulase proteins, useful in detergents and for
 producing 'stonewashed' denim

XX Disclosure; Page 24-25; 65pp; English.

XX The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAV13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel E6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAV13493; Cel E3/B5 is shown in AAV13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new
 CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

XX Sequence 6415 BP; 2067 A; 1082 C; 1689 G; 1541 T; 36 other;

Query Match 10.3%; Score 417.4; DB 20; Length 6415;
 Best Local Similarity 56.3%; Pred. NO. 1.1e-124;
 Matches 893; Conservative 1; Mismatches 642; Indels 51; Gaps 4;

QY 132 tttaccattaaagtcgatacatccaaagcagtcgtgaagcctattagcccttatattacgg 191
 Db 4328 tgttaaaatcatcgatcgatcgtccagggtgaagaaacaaagataaagccgtatattatgg 4387
 QY 192 tagcaatcaggagtggtggcaggcgatgagaatcgtgactgccagcagcttgggtggaatcg 251
 Db 4388 agcaaatcaggatattccagggtgtgtttccacc---ctgcaagacgacttggcgggaacag 4444
 QY 252 aatgaccggatatactgggaaaaacaatatgtccaatgcaggaaagcgtggtgacgtc 311
 Db 4445 attgacgggtttacaattgggagaacaatatgtccaatgcaggaaagcgtggtgacgtc 4504
 QY 312 cagcgatagctattatgcgacaacgcgagtcgataacaaagcgcgaatgtgaaaaaccagg 371
 Db 4505 aagcgatgattatgtgtattattatgggtataacagggaatgataagaacgttccagc 4564
 QY 372 tgcggtggcaacctgtttcacgatacaatcgtgaagcagggcacatatcttttagtcac 431
 Db 4565 agctgtgtgaagcaaatcttcacgagcagtcataaagaacaaatgcatacttcagccatcac 4624
 QY 432 actgcgcagtggcgggttatgtgccaagatggaacacggaagtgtgcaggaaagcgaataa 491
 Db 4625 attacagatggtagggttatgtgccaagatggaatggtacagtgagcgagtcagagac 4684
 QY 492 ggcctcttcgcgtcgggtggaatgagtcgtataaacgttaaaacgttcaatttcatttgc 551
 Db 4685 agctcgcgcgagatggcgtgaggtcaagtttaaaaagatggtgcactgtcttgc 4744
 QY 552 gcctgatctgaaagcaaatcaggtttatgcggtatgaatcgtcaactcttttagtgaataaa 611
 Db 4745 gcctgacgtgaataactatgtatatatgagtgagtttataactatctgtattataa 4804
 QY 612 gtacggcgtgtgcttcaacaaacggcgtgaaaggtactcgtcgcacatgaacccgc 671
 Db 4805 gtatggtcgtatcgtctgcacacgggaattaaaggataataacttgacaacagcagccgga 4864
 QY 672 tctcgtgctgcatacgcacccgcattcatggtgaaaagtcggagcgaagagttggt 731
 Db 4865 cttagtggttactactcactccgcgaattccaccagaaggtaacctcagtgaaattgat 4924
 QY 732 agaccggtcgtgaagtcttaccaaagccgctaagcgggttgacgcgggtgcggaaatttt 791
 Db 4925 aaataaatcgtgtggagcgtggcgaagtaataaagacacttgatccagatgcagaaatttt 4984
 QY 792 tgggcccgttcttaccggtttttggcgcctataaagatcttcaaacctgcacatgattgaa 851
 Db 4985 tggacctcgtatggtttgtgggatatttaacattgcaggtgacactgactgaa 5044
 QY 852 ctctgaaaaaggcaactacagctgttcgttggtgacttaccctcgatcaaatgcgcctcag 911
 Db 5045 tcaggttaaaagaaatacagatggtttttgagctggtcaccttgagcagaagaaagc 5104
 QY 912 ctgcgaagccgaaggaagagattgctggtggtcttcgtatgacactgggtatccctgaagc 971
 Db 5105 atcggtatggtttgggaaraggtttattggtgacttgacatacactgggtaccgcggggc 5164
 QY 972 gatggcggagcagcagcaataataaataagagtgaggcaacgac-----gaaacgaa 1022
 Db 5165 gcaggttggcgggtgtggaataatcttctgacggtgaaatagtacttcaagggatgtggc 5224
 QY 1023 gaaagccagaaatgcaagcgcctcgttactttgtgggtaccacacctacacag----- 1071
 Db 5225 aatgacgaggtgcagccagcagaaagcgtatgggtatccgcacataataaaccacccagaa 5284
 QY 1072 -----gaagtagctgctgcctcaatgggaacagtgaaattcttgc 1112
 Db 5285 aggtcagataacagcgggagaaataagctggataaaccaatggtttccagagtattctcc 5344

QY 672 tctctggtcgatcagcatccgcgattcattcgtgtgaaaggtcgagcgaaagagttggt 731
 Db 4866 cttatggtttactactccgcgattcattcattccacagaggttaacctgcagtgattgat 4925
 QY 732 agaccggtcggttaagtattatccaaagcgcgaagcgttgtaacgcgggtgcggaatttt 791
 Db 4926 aaataaactcggcgagcgtgcgaagagtaataaagacacttgatccagatgcagaaattt 4985
 QY 792 tggcccggtttttacggttttggcgccctataaagatctcaaacctgcacctgattgaa 1851
 Db 4986 tggacctgcattggttttgggatatttaaacattgcaggtgcacctgactgaa 5045
 QY 852 ctctataaagcaactacgctggttgcgttgactattacccatgcatacaaatgcctcga 911
 Db 5046 tcaggttaaaagaaatcacatggttttgcgttgacttgacatacacatggtgcagatgaagaa 5105
 QY 912 ctccgaagccgaagcaagagatgctggtgcttcttgatgatacacatggttgcctcga 971
 Db 5106 atcggatgatttgggaaraggttattggtgacttgacatacacatggtgcagagc 5165
 QY 972 gatggcgaggatcacgaataacaaatgaggtgagcaacgac 1022
 Db 5166 gcaggttgcggtgctgaatgctttgacgtgaaatagtagtacttcaagggtgctgc 5225
 QY 1023 gaaagccagaatgcaagcgcctgacttcttgggtacccgacctacaag 1071
 Db 5226 atagcggagatgcagcaggaagcgtatgggtacccgacatataaaccaccacgaa 5285
 QY 1072 -----gaagatagctgctgctcaatggaacagtggaacagtggaattctgccc 1112
 Db 5286 aggtcagataacagcgggagaaatagctggataaaacaaatggttccagagtatctcc 5345
 QY 1113 ttactgcctcgatataaagcgtcgttggatagattatcccggaacacagctggttt 1172
 Db 5346 actgcttcccaataaaggcagatagacaaagattatctcgtgtaccacacttgctat 5405
 QY 1173 gactgagtatgctggtgcgaaatgatttccggcggtatcgtctatgcccagatgt 1232
 Db 5406 aactgagttgattgagggagggaccatataatcgggaggaatagctttagcagatgt 5465
 QY 1233 gctgggcatcttggcgaacacgcttttatatgccaactactggaagttaaagagatgg 1292
 Db 5466 gttaggatattcggcaagtatgagtatatacagtcagcaagatg -----gggaga 5516
 QY 1293 tgccaacaactacgttagtcggttcacagcgttaccgcaatattatgacggaaagctc 1352
 Db 5517 ttcggggagctatgcacagcggcggtacacacatttatctcaactatgatggaaagttc 5576
 QY 1353 tacttgcgtgatatcagcgttcatgcgcaaacgtcgatattgtaataagctcggtgca 1412
 Db 5577 gagatacggttcaacgtgtgtgagcgtgagacaactgacgttgagaacatgcggtata 5636
 QY 1413 tgcttcggtaacggtgcatctctacaagaactgcaactgctgtgtcatgaataaagcat 1472
 Db 5637 tgcttcaattgaggagaagatgattcgcgtgcatattatatttaataacaggatga 5696
 QY 1473 ggaagtcattgcagcccaattgattcttccgcgagacgacttaacgttccggtgaa 1532
 Db 5697 tgacggaaactgaaggcagagataaagatgaataataccagggttatcacacaggtgaga 5756
 QY 1533 agtatgggttcgacaaaaatagctgcgcaaatgaagagcagcccaatcaccaat 1592
 Db 5757 gatacaggtttgacagtcacagctctcagatcaggaagatgggagtgctcagtaata 5816
 QY 1593 ttacggcaacggtttacattacagctacccgcttttgaagcgtttatcacatggttgac 1652
 Db 5817 acaaaacaacacatacaccatagagttccaaatcgcaggtatacattgtttttaac 5876
 QY 1653 tgcggcgcaatgataccactgtgaaaa 1679
 Db 5877 ttctcaagtagattaaagaataaaa 5903

RESULT 8

AAZ45347

ID AAZ45347 standard; DNA; 2445 BP.

XX AC AAZ45347;

XX DT 27-MAR-2000 (first entry)

XX DE DNA encoding a Bacillus sp. mannanase enzyme.

XX KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 plant material degradation; recycled waste paper; paper making pulp;
 guar; locust bean gum; mannan-containing food; coffee extract;
 cleaning composition; machine washing; hard-surface cleaner;
 dishwashing; oral; dental; contact lens; body-care composition;
 fabric softener; oil well drilling; subterranean formation fracture; ss.

XX OS Bacillus sp.

FH Key Location/Qualifiers

CDS 1..2445

/*tag= a

/product= "mannanase"

/note= "not termination codon given"

FT sig_peptide 1..87

/*tag= b

FT W09964619-A2.

XX 16-DEC-1999.

XX 10-JUN-1999; 99WO-DK00314.
 10-JUN-1998; 98US-0111256.
 20-OCT-1998; 98DK-0001340.
 28-OCT-1998; 98DK-0001341.
 28-OCT-1998; 98US-0105970.
 23-DEC-1998; 98US-0106054.
 05-MAR-1999; 98DK-0001725.
 05-MAR-1999; 99DK-0000306.
 05-MAR-1999; 99DK-0000307.
 05-MAR-1999; 99DK-0000308.
 09-MAR-1999; 99DK-0000309.
 10-MAR-1999; 98US-0123543.
 10-MAR-1999; 98US-0123623.
 11-MAR-1999; 99US-0123641.
 11-MAR-1999; 99US-0123642.

(NOVO) NOVO-NORDISK AS.

PI Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;

WPI; 2000-105891/09.

P-PSDB; AAY54134.

PT New mannanases for treatment of textiles, plant material and coffee
 extract, and in cleaning compositions -

XX Disclosure; Page 230-231; 242pp; English.

XX The present sequence encodes a mannanase enzyme (also known as
 mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
 CC The mannanase hydrolyses galactomannans. Specifically, mannanases
 CC hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
 CC glucomannans, and galactoglucomannans. The mannanase protein, or
 CC preparations containing it, are used to improve properties of cellulosic
 CC or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
 CC sizes or printing pastes). They are also used to degrade or modify
 CC plant materials (particularly recycled waste paper, paper making pulps,

CC or material containing guar or locust bean gums (thickeners), or to
CC reduce viscosity of mannan-containing foods (or feeds). The mannanses
CC are also used to process coffee extracts (to inhibit gel formation); in
CC cleaning compositions (for machine washing of fabrics, as hard-surface
CC cleaners, for hand or machine dishwashing, also in oral, dental, contact
CC lens or body-care compositions) where they remove mannan-containing
CC soils and prevent binding of some soils to celluloses; and in fabric
CC softeners. They can also be used in oil well drilling to fracture
CC subterranean formations.
XX
SQ Sequence 2445 BP; 683 A; 535 C; 642 G; 585 T; 0 other;

Query Match 9.8%; Score 399.2; DB 21; Length 2445;
Best Local Similarity 60.4%; Pred. No. 5.3e-119;
Matches 713; Conservative 0; Mismatches 458; Indels :9; Gaps 3;

QY 2126 ctctcatgatgacaaacacgacgaatttctttaaagtgcaaaagctgtgagagctgg 2185
DB 707 ctcttatggcgacaaacaaataatctgacagtgaatgacaggttacccgtcaatcttg 766
QY 2186 tacttggaaaacgctgattttaaaagaaacgtccggtgtgaagttctgttgaaatgcaag 2245
DB 767 acttgaagagacagaagtctctcgtggtgaattgaatgctgcgcatcgttaagtctcaatgaag 826
QY 2246 gggcgaatcagatcggttttgaacagcgctgggctggtatgatcatcagactcgtcaagc 2305
DB 827 gcgaaacacataacacccatcctatgtggttgggagtggtacataatcattatatacagc 886
QY 2306 tggagccag-----ctgctgaccgtccaccgcatcggtgaacgaaacgctcatcaatc 2359
DB 887 ttgtacctgtgtcagttcgtatccgacccgacacgcatcaggtcgaaacacactgtggaatc 946
QY 2360 cgaatgcgacagtagaagcaaaagcatgatgaactacctggttgaatcaatacgggaaga 2419
DB 947 cggacccctcaactgaggaagcgttaataatattctctagacagtagcagtcgggaaca 1006
QY 2420 atatgtttccggtcaagaggtatgccgaaattgattggtcttcaacgcaatgtagta 2479
DB 1007 aaattctacaggtcaaaacccaggttgaagacgcccaggtggatccatgaacagggcgga 1066
QY 2480 aaaaacgctgattgctggcactgacctgattgactattcccccaacgagagcggaacacg 2539
DB 1067 aatatectcggttattgaggttattttatgactacagccgctcccgctgagtgcatg 1126
QY 2540 gtcttagttccacagacggaagagcgattgaatggatggatcaagaggggcatgttta 2599
DB 1127 ggcgaactggaactcggttggaggaagcagattgagtggaagagagtggtggatcatta 1186
QY 2600 ctttgcattggaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 2659
DB 1187 ctttgcattggaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1246
QY 2660 ggagaggttctatgcgacttgcactacattgcataatgcataatgcataatgcataatgc 2719
DB 1247 ggtccggttttataccgtgccaacacgttggatggatggatggatggatggatggatgg 1306
QY 2720 agtccgaagattataaattgcttattcgcacatgcacatgcacatgcacatgcacatgcac 2779
DB 1307 aactgagatttccaaattggtattgtagcagatggatggatggatggatggatggatggat 1366
QY 2780 agtgcagagatcgaaagtctctgctctctctctctctctctctctctctctctctctctct 2839
DB 1367 ggcctgagcagagaacacccctgtgtatggagaccgtctctctctctctctctctctct 1426
QY 2840 ggtctggtgggagcgttaaagtcctgagcctgtttaaataattgtattttgtatgacgacg 2899
DB 1427 ggttctggtgggagcgttaaagtcctgagcctgtttaaataattgtattttgtatgacg 1486
QY 2900 atcgtttgactaatgtgcacaaatgaacaaatctgatctggttgggaactcgttctgtc 2959
DB 1487 atcgtttacccaatacccaataacaaatctgatctggttgggaactcgttctgtc 1546

RESULT 9

AAZ45348
ID AAZ45348 standard; DNA; 1488 BP.
XX AC AAZ45348;
XX AC AAZ45348;
XX DT 27-MAR-2000 (first entry)
XX DE DNA encoding a Bacillus sp. mannase enzyme.
XX DE Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW plant material degradation; recycled waste paper; paper making pulp;
KW guar; locust bean gum; mannan-containing food; coffee extract;
KW cleaning composition; machine washing; hard-surface cleaner;
KW dishwashing; oral; dental; contact lens; body-care composition;
KW fabric softener; oil well drilling; subterranean formation fracture; ss.
XX OS Bacillus sp.
XX FH Key
XX CDS Location/Qualifiers
FT 1..1488
FT /*tag= a
FT /product= "mannanase"
FT /note= "not termination codon given"
FT sig_peptide 1..111
FT /*tag= b
XX PN WO9964619-A2.
XX PD 16-DEC-1999.
XX PF 10-JUN-1999; 99WO-DK00314.
XX PR 10-JUN-1998; 98US-0111256.
XX PR 20-OCT-1998; 98DK-0001340.
XX PR 20-OCT-1998; 98DK-0001341.
XX PR 28-OCT-1998; 98US-0105970.
XX PR 28-OCT-1998; 98US-0106054.
XX PR 23-DEC-1998; 98DK-0001725.
XX PR 05-MAR-1999; 99DK-0000306.
XX PR 05-MAR-1999; 99DK-0000307.
XX PR 05-MAR-1999; 99DK-0000308.
XX PR 05-MAR-1999; 99DK-0000309.
XX PR 09-MAR-1999; 99US-0123543.
XX PR 10-MAR-1999; 99US-0123623.

mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre; synthetic fibre; yarn; fabric; printing paste; thickener; viscosity; plant material degradation; recycled waste paper; paper making pulp; guar; locust bean gum; mannan-containing food; coffee extract; cleaning composition; machine washing; hard-surface cleaner; dishwashing; oral; dental; contact lens; body-care composition; fabric softener; oil well drilling; subterranean formation fracture; ss.

Humicola insolens.

Key Location/Qualifiers
CDS 1..1464
/*tag= a
/product= "mannanase"
/note= "not termination codon given"
sig_peptide 1..63
/*tag= b

W09964619-A2.

16-DEC-1999.

10-JUN-1999; 99WO-DK00314.

10-JUN-1998; 98US-0111256.

20-OCT-1998; 98DK-0001340.

20-OCT-1998; 98DK-0001341.

28-OCT-1998; 98US-0105970.

28-OCT-1998; 98US-0106054.

23-DEC-1998; 98DK-0001725.

05-MAR-1999; 99DK-0000306.

05-MAR-1999; 99DK-0000307.

05-MAR-1999; 99DK-0000308.

05-MAR-1999; 99DK-0000309.

09-MAR-1999; 99US-0123543.

10-MAR-1999; 99US-0123623.

10-MAR-1999; 99US-0123641.

11-MAR-1999; 99US-0123642.

(NOVO) NOVO-NORDISK AS.

Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad MS;

WPI; 2000-105891/09.

P-PSDB; AAY54128.

New mannanases for treatment of textiles, plant material and coffee

extract, and in cleaning compositions

Disclosure; Page 221; 242pp; English.

The present sequence encodes a mannanase enzyme (also known as mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase). The mannanase hydrolyses galactomannans. Specifically, mannanases hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans. The mannanase protein, or preparations containing it, are used to improve properties of cellulosic or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based sizes or printing pastes). They are also used to degrade or modify plant materials (particularly recycled waste paper, paper making pulps, or material containing guar or locust bean gums (thickeners), or to reduce viscosity of mannan-containing foods or feeds). The mannanases are also used to process coffee extracts (to inhibit gel formation); in cleaning compositions (for machine washing of fabrics, as hard-surface cleaners, for hand or machine dishwashing, also in oral, dental, contact lens or body-care compositions) where they remove mannan-containing soils and prevent binding of some soils to celluloses; and in fabric softeners. They can also be used in oil well drilling to fracture subterranean formations.

Sequence 1464 BP; 290 A; 448 C; 456 G; 270 T; 0 other;

Query Match 8.1%; Score 327.2; DB 21; Length 1464;
Best Local Similarity 54.3%; Pred. No. 1.3e-95;
Matches 728; Conservative 0; Mismatches 603; Indels 9; Gaps 3;
QY 1931 cgcgggcccctgtgacggtccgagtcgctatgaagcgggaagatggcagcgtgaaggga 1990
Db 116 ctccgctccctgtgcccgggtccgagaccttcgaagcggaggtgccatccatccagggca 175
QY 1991 ccgtgttggaatccagtgggaccggtctccggtacttggttaactaactatttcacaca 2050
Db 176 cgagggttgagtcgagcctcgccggtactctgtaccggatagtacgggttcgacg 235
QY 2051 atgcaggggattccctgacgatgcagatccagggtcccccagcggcgggtttgacaaattta 2110
Db 236 agccagtgacaagatcacgttccacgtggacagcagagaccacacggtgtacgacctca 295
QY 2111 caatcgatataccgttctctctctatgatgacaaacgcagaaatttctttaaattgcaaaag 2170
Db 296 ccatcccggtgcccgcctctatggagaaagcgcacacccgctgctcaataacggcg 355
QY 2171 cgtctggagagctggtacttttgaaaacggcgtgattttaagaaacgtcccggtggttaag 2230
Db 356 cggcaagtggaggtctacttcccgccagcgattctgtcgacatcgctgccggccagg 415
QY 2231 ttctgttgaaatgcaggggcgaatacagatcggttttgaaacaggcgtggggtgtgatgata 2290
Db 416 tccgtgtaaccaggcggcgaacacacatcgacattgtcaacaactggggtgtgtacctga 475
QY 2291 tcgactacgtcaagctggagccagctgctgacccgtccacgcgctgcggtgaacgaaaaacgc 2350
Db 476 tcgactcacaatcacatcacccccctcccgccgcgcacccccctcacaatacaacccctccc 535
QY 2351 tcataatccgaatgcgacagtagaagcaaaagcattgtgaactacactggttgatcaat 2410
Db 536 ccgtcaacccctgcgcgcgacgacaaacgcggcgtgtgtacgacatacctcgctccatct 595
QY 2411 acgggaagaataatcttccggtcgaagagatagccgaaattgattggttcacagcga 2470
Db 596 acggcaagaataatcttccggccagcagagcttcttggcggaactggtgacgccaac 655
QY 2471 atgtagtgtaaaagccggtctattgcgcaacttgacctgattgactatcccccaagcagag 2530
Db 656 agacgggcaaaaccccgctggtgctccgtcgatagtgattatccccctagtctgg 715
QY 2531 cggaaacacggtcttagttccacagagacggaagcggatgaatgggataaagcagggg 2590
Db 716 tggaaagagcactgctggtgctgctgagagggcactcgagcatcacgcggcgcg 775
QY 2591 gcattgttactttgcatggcactggaacgcgcgaaagctctgacgacacacgagga 2650
Db 776 gcattgtctcgggtgtgtgcaactggaacgcgcacggggctgtacgacacacgagc 835
QY 2651 aagaatggtgagagggttctatgcccgtattcgactacattcgatagataagatgacgag 2710
Db 836 gccggtggtgagcgggttctacacgacgacgacgactttgacgctgcgcgcgctgg 895
QY 2711 atcatccagatccgaagattataattgtttattgacacatcgatgattgacgggc 2770
Db 896 cgatacgcgaatgccaactacacgctgctgacccgggataccgacgacgacgctgcggtgc 955
QY 2771 aattgaagaagtgcagatgcaaaagtctctgcttcttcttcccttcttcacgagcgg 2830
Db 956 agctcaagaggttcgggacgcggcggtgcggtgtgttggcccgctgcgacgagccg 1015
QY 2831 agggcaaatggttctggtggggcgttaagggtcctgagcctgttaaaaaatttatatt 2890
Db 1016 agggcgtgtgttgggtgggagcgaagggcccgagggcgcatacaagaagctgtggggga 1075
QY 2891 tgaatgcacatcgtttgactaattgtcacaaattgacaaatgacatctgctgggtctggaact 2950
Db 1076 tctgtatgaccgactcaagaaactaccatgggtgctgaatacctgctgtgtgtggaact 1135
QY 2951 ctgttctcccgactggttatcccgaggatgagtagtattgagcttgcactctt 3010

[illegible]

RESULT 11

AAZ45350

AAZ45350
ID AAZ45350 standard: DNA: 3041 BP.[illegible]

AAZ45350;

XXXXXX

DT	27-MAR-2000	(first entry)
XX		
DE		
XX		
XX		DNA encoding a mannanase enzyme.
XX		
KW		Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
KW		endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
KW		mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
KW		synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
KW		plant material degradation; recycled waste paper; paper making pulp;
KW		guar; locust bean gum; mannan-containing food; coffee extract;
KW		cleaning composition; machine washing; hard-surface cleaner;
KW		dishwashing; oral; dental; contact lens; body-care composition;
KW		fabric softener; oil well drilling; subterranean formation fracture; ss.

Caldocellulosiruptor sp.

Accession	Protein	Location/Qualifiers
CS	Carboxypeptidase	
XX		
FH	Key	
FT	CDS	292..3001
FT		/tag= a
FT		/product= "mannanase"
FT	sig_peptide	1..66
FT		/tag= b

AA
PN
W09964619-A2-

PN W03904019-AZ
XX

16-DEC-1999

XX
TO
DE
TO
XX

PF 10-JUN-1999; 99WO-DK00314.

XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700

PR 10-JUN-1998; 98US-0111256. PR

PR 20-OCT-1998; 98DK-0001340.

PR 20-OCT-1998; 98DK-0001341.

PR 28-OCT-1998; 98US-0105970.

PR 28-OCT-1998; 98US-0106054.

PR 23-DEC-1998; 98DK-0001725.

PR 05-MAR-1999; 99DK-0000306.

(NOVO) NOVO-NORDISK AS.
Kauppinen MS, Schuelein M, Schnorr K, Andersen LN, Bjornvad ME;
WPI: 2000-105891/09.
P-PSDB; AAY54137.
New mannanases for treatment of textiles, plant material and coffee
extract, and in cleaning compositions -
Disclosure: Page 237-238; 243pp; English.
The present sequence encodes a mannanase enzyme (also known as
mannan endo-1,4-beta-mannosidase, beta-mannanase, or endo-1,4-mannanase).
The mannanase hydrolyses galactomannans. Specifically, mannanases
hydrolyse 1,4-beta-D-mannosidic linkages in mannans, galactomannans,
glucomannans, and galactoglucomannans. The mannanase protein, or
preparations containing it, are used to improve properties of cellulosic
or synthetic fibres, yarn or (non)woven fabrics (removal of mannan-based
size or printing pastes). They are also used to degrade or modify
plant materials (particularly recycled waste paper, paper making pulps,
or material containing guar or locust bean gums (thickeners), or to
reduce viscosity of mannan-containing foods or feeds). The mannanases
are also used to process coffee extracts (to inhibit gel formation); in
cleaning compositions (for machine washing of fabrics, as hard-surface
cleaners, for hand or machine dishwashing, also in oral, dental, contact
lens or body-care compositions) where they remove mannan-containing
soils and prevent binding of some soils to cellulosics; and in fabric
softeners. They can also be used in oil well drilling to fracture
subterranean formations.
Sequence 3041 BP: 1006 A; 518 C; 728 G; 789 T; 0 other;

PR 10-MAR-1999; 99US-0123641.
 PR 11-MAR-1999; 99US-0123642

Db 2379 caggctgaaggtgtgccagttcttttcagaccgcttcacagagcctctggtggttc 2438
 QY 2845 tgggtggcgctaaaggtctgagcctgttaaaaaattgtatatttggatgcacgatcgt 2904
 Db 2439 tgggtgggtgcaaaaggtccagagcgtatataaaagctttggaagctcatgtttgacagg 2498
 QY 2905 ttgactaatgtgcacaaattgacaaatcgtatcgtggtctggaactctgttgcctccgac 2964
 Db 2499 cttgtaaatcatcacaaatacaaaactaataatgggtatggaaacggtcagagatgctgcc 2558
 QY 2965 tgggtatccgggagatgagtatgttgatattttgagcttcgactcttattccgcaagcagc 3024
 Db 2559 tgggtatccgggtgacagttgtgatataattgcagaatatatatgaggaagaaagct 2618
 QY 3025 gactacagccgcgaattgcaaaatgaagacctgtttacatttgggcaagggacaaaag 3084
 Db 2619 cagtaactaccatatacagagaggttctgaaagctctca-agtacaaatgcaaaaca 2677
 QY 3085 ctgatttgcctgagcgaacacgacctatcccgaccggtatctgtatgaaagcgtatca 3144
 Db 2678 gatgatagcactttctgagtcggaactattctgacccggtctgtctaaacaagaagg 2737
 QY 3145 agccattgagctggttgcctacatgtatg---gggatttcttgagagcggcaaaaca 3201
 Db 2738 tgttctggtggtgtgtttctgtatggcaggaagctatgtcatgacagcgagcaagta 2797
 QY 3202 aaacgtcctctggagc 3217
 Db 2798 caacgatgaatggaac 2813

RESULT 12

AAQ15178
 ID AAQ15178 standard; DNA; 2600 BP.

AC AAQ15178;

17-MAR-1992 (first entry)

Portion of pAEC-1 contg. cellulase AE-1 gene.

Detergents; pharmaceuticals; deinking; carboxymethylcellulose; ss.

Aeromonas strain no. 212.

Key Location/Qualifiers
 CDS 192..2540
 /*tag= a
 /product= cellulase AE-1

JP03251174-A.

08-NOV-1991.

28-FEB-1990; 90JP-0045465.

28-FEB-1990; 90JP-0045465.

(OJTP) OJI PAPER KK.

WPI; 1991-373412/51.

P-PSDB; AAR15625.

Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -

of opt. pH when carboxymethylcellulose is used as substrate.

Claim 3; Fig 2; 8pp; Japanese.

The plasmid, pAEC 1, was prepd, ligating chromosomal DNA contg. the
 gene (obtd. from Aeromonas) into pUC18. The gene can be used to
 produce recombinant enzyme which is used for the effective utilis-
 ation of biomass resources and the mfr. of pharmaceuticals and
 foodstuffs, and also for the detergent and deinking of waste paper.

XX Sequence 2600 BP; 687 A; 649 C; 700 G; 564 T; 0 other;
 SQ

Query Match 3.3%; Score 133.2; DB 12; Length 2600;
 Best Local Similarity 56.0%; Pred. No. 5.9e-32;
 Matches 252; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 3594 gggattgtgtcctcagctatcgacagcggtatgacaaatgtgaaagcacaatcacttgaacc 3653

Db 2087 ggtattggcgttcagtaaaaaacggagatacgaatgcagcgcacacacagttcaagcc 2146

QY 3654 gcatttccaaatttaataaagttacaattcctgaccgatcaacagagttgaaaaattcg 3713

Db 2147 tcatttcaatatcgtgaacaaagggggtggtggtggtggtggtggtggtggtggtggtg 2206

QY 3714 ctactactacacgtacggtgacgtgagcagacacattcaactgcgactatgcggtgct 3773

Db 2207 atattattcacagctgacggcaatgaccaactgcaataattgtgactgggcatggt 2266

QY 3774 gagctgtcgaagctgaatggttaagctggttaaaatgataaagctgaaccggtgctga 3833

Db 2267 aggtgtctcaacctgaacggggttctgtgaaatgaatccgggcaagcgaacgcga 2326

QY 3834 ttattttggaactcagcttcaactcgatgcagcgtgttagctctctgaggaagcac 3893

Db 2327 tacctacttgagatcacgtttaaagccgagcgggtctgctgcagcccggtgcaaacg 2386

QY 3894 gggcggaattcaactcgtatttcaaaagcagactggtcgaaactataacgaagtgacga 3953

Db 2387 ggggtatcacagcgcgaatcatcgtcggaactggcgcaactttaacgaagcaatga 2446

QY 3954 ttactgtataaagtcagcagacttcattcgcagatcatcacgaagactcgttgtatca 4013

Db 2447 ctattctttgatcccaacgaacgcttatgcaaatgggaaacgctgcacctgtacca 2506

QY 4014 caatggcgtactgtttggggaacccgaacc 4043

Db 2507 taacggtacactggtgtctcggaacgagcc 2536

RESULT 13

AA55661

ID AAX55661 standard; DNA; 11706 BP.

AC AAX55661;

30-JUL-1999 (first entry)

DNA sequence encoding truncated cellulases Cel B4/5 and Cel B5.

Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

cotton-containing fabric; stonewashing; ss.

Unidentified.

EP921188-A2.

09-JUN-1999.

15-SEP-1998; 98EP-0810919.

19-SEP-1997; 97US-0932571.

(CLRN) CLARIANT FINANCE BVI LTD.

Anderson P, Bergquist PL, Daniels RM, Farrington GK;

Gibbs MD, Morgan H, Williams DP;

WPI; 1999-315403/27.

P-PSDB; AAV13492.

QY 3767 cggctgctgagctgctgaagctgaatggttaagctggtttaaataatgataaagctgaaccg 3826
 Db 4210 cacagatagggcgaagcaatgacattcaatttgaagctgagcgcggagtgagtg 4269
 QY 3827 gtgctgattattattggaagtcagcttcaactcgagtgagcggctgttagcctctgag 3886
 Db 4270 gagcggattattacttggaggttagatttagcagtgagctggcagcttgta 4329
 QY 3887 gaagcacggcggaattcaactctgattcataaagcagactggctgcaactataacgaaa 3946
 Db 4330 aggacacaggggatacacaggtaaggttttaacaagaatgactggagcaattacaatcagg 4389
 QY 3947 gtgagcattactcgtataaaggtacgagacttcttcaagcagatcatcacgaaagctact 4006
 Db 4390 cagagactggtcaggttgcagatgacgacgaattatggagagaatgcgaagtgtaacgc 4449
 QY 4007 tttatcaaatggcgtacttctgttggggaaccgaaaccg 4044
 Db 4450 tttatgtagatggttctgttggatgggagcggagcgg 4487

RESULT 15

AAQ49820

ID AAQ49820 standard; DNA; 2175 BP.

AC AAQ49820;

DT 27-APR-1994 (first entry)

DE NK-1 cellulase gene.

KW Cellulase; pH dependence; mutation; ss.

OS Bacillus N4.

FH Key Location/Qualifiers

FT CDS 438..1937

FT /*tag= a

FT /product= cellulase_NK-1

PN JP05236969-A.

XX 17-SEP-1993.

XX 28-FEB-1992; 92JP-0075883.

XX 28-FEB-1992; 92JP-0075883.

XX (BEPP/) BEPPU T.

XX WPI; 1993-330585/42.

XX P-PSDB; AAR42122.

Changing the pH-dependence of cellulase enzymatic activity - by changing base sequence of cellulase-producing gene of Bacillus microbe to base sequence coding asparagine and serine residues at specified aminoacid sites

Claim 1; Fig 1; 9pp; Japanese.

Sequences (AAQ49818-19) consist of two synthetic fragments which are used to induce a mutation within the cellulase NK-1 gene. The gene is shown in sequence (AAQ49820). The modified NK-1 gene shows a change in pH dependence.

Sequence 2175 BP; 715 A; 434 C; 496 G; 530 T; 0 other;

Query Match

Best Local Similarity 53.2%; Score 87; DB 14; Length 2175;

Matches 258; Conservative 0; Mismatches 215; Indels 12; Gaps 3;

QY 3559 gcagcgcgaagcggatattactcagcagccgacacgggattggtgtccagtatcgaca 3618

Db 1458 gaacgcgcgcacagaataaccgcgcaggaacggcatttctgtacaatacaagca 1517
 QY 3619 gcggatacaaatgtgaacgacaatcacttgaacccgcatcttccaaatttttaataaagg 3678
 Db 1518 ggggatgggggtgtgaacagcaaccaaatccgcgcgagcttcacataaaaaataacggc 1577
 QY 3679 acaatctccgtaccgatcaacgaggtgaaaattogctactactacacgactcgacggtgac 3738
 Db 1578 aatgcgacggttgattttaaagatgtcactgccgtttactgggtataa---cgcgaaaaac 1634
 QY 3739 cgtgagcagacattcaactgcgactatgcggtgtgctgagctgctcgagctgaatggttaag 3798
 Db 1635 aaggggccaaacttggactgtgactacgcgagattggatgcggcaatctgaccccaaa 1694
 QY 3799 ctggttaaaatggataaagctgcaacggctgctgattatttttggaaagtcagcttcaac 3858
 Db 1695 ttgtgacgctgcataaacctaaagcaaggtgcagataacctatctggaactgggtttta-- 1752
 QY 3859 tcggatgcagcgtgttagcctctggaggaagcagcagggcggaattcaaaactcgtattcat 3918
 Db 1753 ----aaacaggaacgctgtcaccggggagcaagcaggggaatattcagcttctcttcac 1808
 QY 3919 aaagcagactggtcgaaactataacgaaagtgcgattactctgtataaaaggtacgcagact 3978
 Db 1809 aatgatgactggagtaattatgcacaaagcgcgattattccttt---tttcaatcaaat 1865
 QY 3979 tcattcgacgatcatcacgaagctagttgtatcaaatggcgtacttgtttggggaacc 4038
 Db 1866 acgtttaaacacgcaaaaaattacattatatacatcaaggaactgatttgggggaaca 1925
 QY 4039 gaacc 4043
 Db 1926 gaacc 1930

Search completed: July 12, 2002, 19:12:29

Job time: 31626 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 13, 2002, 00:46:41 ; Search time 261.94 Seconds
(without alignments)
3803.503 Million cell updates/sec

Title: US-09-784-554B-3

Perfect score: 4056

Sequence: 1 atgaagcgcaaaatagtag.....ccgaaccgacagctaattaa 4056

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	630.2	15.5	2977	2	US-07-862-588B-1
2	629.4	15.5	1854	1	US-08-361-920-28
3	629.4	15.5	1854	1	US-08-479-339-28
4	629.4	15.5	1854	1	US-08-483-432-28
5	417.4	10.3	6416	4	US-09-136-574A-2
6	90	2.2	11707	4	US-09-136-574A-1
7	85.2	2.1	2029	4	US-09-136-574A-46
8	63	1.6	1624	2	US-07-862-588B-6
9	63	1.6	1775	2	US-07-862-588B-5
10	51	1.3	7218	1	US-08-232-463-14
11	45.6	1.1	1011	3	US-09-095-163-1
12	35.8	0.9	1146	4	US-09-277-716-21
13	34.8	0.9	289	4	US-09-007-005-17
14	34.8	0.9	289	4	US-09-244-796-17
15	34.2	0.8	486	1	US-08-048-164A-1
16	34.2	0.8	486	1	US-08-048-164A-3
17	34.2	0.8	486	1	US-08-460-462-1
18	34.2	0.8	486	1	US-08-460-462-3
19	34.2	0.8	486	1	US-08-460-457-1
20	34.2	0.8	486	1	US-08-460-457-3
21	34.2	0.8	486	1	US-08-460-458-1
22	34.2	0.8	486	1	US-08-460-458-3
23	34.2	0.8	486	2	US-08-460-455-1
24	34.2	0.8	486	2	US-08-460-455-3
25	34.2	0.8	486	2	US-08-330-394A-1
26	34.2	0.8	486	2	US-08-330-394A-3
27	34.2	0.8	499	3	US-09-006-636-5

28 34.2 0.8 499 4 US-09-006-632-5 Sequence 5, Appli
29 33.4 0.8 1482 4 US-09-198-956-9 Sequence 9, Appli
30 33.4 0.8 1482 4 US-09-198-955A-11 Sequence 11, Appli
31 33.2 0.8 1960 1 US-08-415-823-3 Sequence 3, Appli
32 33.2 0.8 1960 2 US-09-086-662-3 Sequence 3, Appli
33 32.2 0.8 3622 2 US-08-591-079-7 Sequence 7, Appli
34 32.2 0.8 5393 2 US-08-591-079-9 Sequence 9, Appli
35 31.6 0.8 3416 2 US-08-701-240-3 Sequence 3, Appli
36 31.6 0.8 3416 3 US-09-138-236-3 Sequence 3, Appli
37 31.4 0.8 2103 2 US-08-897-340-2 Sequence 2, Appli
38 31.4 0.8 2103 3 US-09-252-329-2 Sequence 2, Appli
39 31.4 0.8 2106 3 US-08-613-009A-4 Sequence 4, Appli
40 31.4 0.8 2247 3 US-08-613-009A-3 Sequence 3, Appli
41 31 0.8 864 4 US-08-858-207A-119 Sequence 119, App
42 31 0.8 4411529 4 US-09-103-840A-1 Sequence 1, Appli
43 30.8 0.8 1170 3 US-08-689-421-20 Sequence 20, Appli
44 30.8 0.8 1170 4 US-09-389-528-20 Sequence 20, Appli
45 30.8 0.8 1170 4 US-09-181-827A-20 Sequence 20, Appli

ALIGNMENTS

RESULT 1
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:

NAME/KEY: CDS
LOCATION: 677..2776
OTHER INFORMATION:
US-07-862-588B-1

Query Match 15.5%; Score 630.2; DB 2; Length 2977;
Best Local Similarity 63.6%; Pred. No. 1.4e-200;
Matches 991; Conservative 1; Mismatches 559; Indels 6; Gaps 2;

QY 120 aacggcaaaagactgtaccattaaagtcgatacatccaagatcgtgaagcctattagccc 179
DB 769 AGCTGCGAGTATGCTACTTTCAGTATATACGAGTTCGAGCTGCGAGCTGACGCC 828
QY 180 ttattattagctaacatcaggaggttggcaggcgtatgagaatctgactgccagacat 239
DB 829 CAATATTTACGGAACCAATCAGGATCTGAGCGGAGGAGAACTGGTTCATCCCGCAGGCT 888
QY 240 tgggtgcaatcgaatgaccggtataactgggaaacaatatatccaatgcaggaagcga 299
DB 889 CGGAGGCAACGGCTGACGGGTTCACACTGGGAGAACACGCAATCCAGCGCCGGAAAGGGA 948
QY 300 ctggatgcagtcagcagatagctatttctgcgcaaacgcggatgtgacaaaagccgaatg 359
DB 949 CTGCTTTCATTACAGCGATGATTTCTCTCGCGCAACGGGTGGTTCAGACACCGACATG 1008
QY 360 tgaagacccaggtgctgggtgcaacctgtttcacgatacaatcgtgaagcaggagacata 419
DB 1009 CGAACAGCCGGGGCGGTGTTTACCGCTTTTTCACGATAAATCTTTGGAGAATGGAGCTTA 1068
QY 420 ttctttagtcacactgcggtgctccgtggtggaatgaggtgtgaaacgctaaaaatcgcc 479
DB 1069 CTCATTTGTACCGTGAATGCGGGGTATGTCCCGGGATTAAGAACGGTCCAGTTGA 1128
QY 480 ggaagcgaaaagctcctccgtggtggaatgaggtgtgaaacgctaaaaatcgcc 539
DB 1129 CGAGAGTGAGACGGCTCCGTCACCGGTGGGATAAGGTGAGTTCGCCAAAATCGCC 1188
QY 540 gttcaattcagcagctgatacgaagacatcagggtttatcgagatgaattcgtaactt 599
DB 1189 GTTCTCCCTTCAGCCTGATCTGAACGACGGACAAAGTGTATATGGATGAAGATTAACCT 1248
QY 600 tttagtgaagaaagtcagcgttcttcaacaaacggcgtgaaagagatactcgtcga 659
DB 1249 CTTGGTCAACCGGTATGGAACCGCTTCAAGCTCAACGGGATCAAAAGCTATTCGCTGGA 1308
QY 660 caatgaacccgctctgtgtcgcatacgcacccgcgcatcattgtgaaaggtcgagc 719
DB 1309 TAACGAGCCGGCGCTGTGTGTGAGACGCATCCAAAGGATTCATCCGAGCAGTTACAAGC 1368
QY 720 gaaagagttgtagaccgctcggtgaagtttaccgaagcgcctaaagcgggttgacgagg 779
DB 1369 GGCAGAACTCGTCTGAAGGACATCGACTTGTCAAAAGGGGGTGAAGAAGCTCGATCCGA 1428
QY 780 tgggaaattttgggcccgttcttcaagctgttggcgcctataaagatcttcaactgc 839
DB 1429 TGCGGAATATTCGGTCTCCCTTTACGGTTTCGGCGCATATTTCTCTCGAGGACGC 1488
QY 840 acctgattggaactctgtaaaggaactacagctggttctgtgactattacactgatca 899
DB 1489 ACCGGATTGGCCGATTGGAAGGCAACTACAGCTGGTTTATCGATTACTATCTGGATCA 1548
QY 900 aatgcgctcagctcgcaagcgaagcgaagatgctggtgtcttcgtgatacactg 959
DB 1549 GATGAAGATGCTCATACGAGAACGGCAAAAGATTGCTCGATGTCTGGAGCTCCACTG 1608
QY 960 gtactctgaagcagatggcgaggcgatacgaattacaatga---ggtaggcaacga 1016
DB 1609 GTATCCGGAGACAGCGGCGGAGCCAGCAATCGTCTTTGGCGGGCGCGCAATATCA 1668
QY 1017 aacgaagaaagcgaatgcgaagcctcgtacttctgtggatccgacacataaggaaga 1076
DB 1669 TAGCGAAGAGGCTCGGTGTACAGCGCAAGATCGCTATGGGATCCGCTTACAGGAAGA 1728

RESULT 2

US-08-361-920-28

; Sequence 28, Application US/08361920

; Patent No. 5457046

; GENERAL INFORMATION:

; APPLICANT: Woelidike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESS: No. 5457046 No. 5457046disk of No. 5457046th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,920

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

APPLICATION NUMBER: DK 1158/90
 FILING DATE: 09-MAY-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/DK91/00124
 FILING DATE: 08-MAY-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Lambiris, Elias J.
 REGISTRATION NUMBER: 33,728
 REFERENCE/DOCKET NUMBER: 3435.204-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-867-0123
 TELEFAX: 212-867-0298
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1854 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHEICAL: NO
 ORGANISM: *Bacillus lautus*/Hemicola insolens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1851
 US-08-361-920-28

Query Match 15.5%; Score 629.4; DB 1; Length 1854;
 Best Local Similarity 63.8%; Pred. No. 1.9e-200;
 Matches 986; Conservative 1; Mismatches 552; Indels 6; Gaps 2;

QY 132 tgttaccattaaagtcgataccaaagcgtgtaagcctattagccttattattacgg 191
 DB 81 TGTCATCTTCACGATTAAATACCGAGTCGGAGCGTCGAGCGATCAGCCCAATATTACGG 140
 QY 192 tacgaatcaggaggtggcaggcgatgagaaatctgactgacgacgactggtggcaatcg 251
 DB 141 AACCAATCAGGATCTCAGCGGGAGCGGAGAACTGGTCATCCCGCAGCTCGGAGGCAACGG 200
 QY 252 aatgacggatataactgggaaacaaatattgccaatgaggaaagcgtgagtgatcagtc 311
 DB 201 GCTGACGGGTTCACACTGGGAGAACAAACGATCCAGCGCGGAGGAGGACTGGCTTCATTA 260
 QY 312 cagcagtagctattatgcgacaacccggtgattgacaaagcgaatgtgaaagccagg 371
 DB 261 CACGATGATTTTCTCTGGGCAACGGTGTGTTCCAGACACCCGACTGGCAGAACCGGG 320
 QY 372 tgcgtggcaacctctttcagatcaatcgctgaagcaggcgacatatcttttagtcac 431
 DB 321 GCGGTTGTTACCGCTTTTACGATAAATCTTTGGAGATGGAGCTTACTCCATTTGTAAC 380
 QY 432 actgacgagtgccggttatgtggcgaagcgaatggaacggaagtgtgcaagaaagcga 491
 DB 381 GCTGCAAAATGGCGGGTTATGTCTCCGGGATAAGAACCGTCCAGTGTGACGAGAGTGAGAC 440
 QY 492 ggtctctcgtcgtggtggaatgaggtcgttaaacgctaaacgttaaaatgcgcgtttcaattgca 551
 DB 441 GGCTCCGTCACCGCTGGGTGATAGGTGCTAGTTGCGCAAAATGCGCGCTTCTCCCTTCA 500
 QY 552 gctgactgaaagacaatcagggttatgctggaatcgtcaacttttttagtgaaaaa 611
 DB 501 GCCTGATCTGAACGAGCGGACAAAGTGTATATGGATGAAGAAGTTAACTTCTTGTCACCG 560
 QY 612 gtacggcgttgttcaacaaacagggcgtgaaagatgactcgtcgcacaaatgaaccgcg 671
 DB 561 GTATGAAACCGCTTCAACGTCACCGGCATCAACGCTATTCCGCTGATTAACGAGCGCGC 620
 QY 672 tctctgtcgtacacgctccggtcattcattgtgaaaggtcgagcgaagagattggt 731
 DB 621 GCTGTGCTGACGACGATCCCAAGGATTCATCCGGAGCAGTTACAAAGCGGACAGACTCGT 680
 QY 732 agaccggtcgttaagtttatccaaagccgctaaggcgttgacggttgagcggggtgagaaatttt 791

DB 681 CGCTAAGAGCATCGACTTTGTCAAAAGCGGTGAAGAACATCGATCGCATGCCGAATATT 740
 QY 792 tgggcccgttcttacctggttttggcgccctataaagatcttcaaaactgacactgattgaa 851
 DB 741 CGGTCTCTGCCCTTTACGGTTTCGGCGCATATTTGCTCTGAGGACGACCGGATTTGCC 800
 QY 852 ctctgtaaaaagcgaactacagctggttgcgtggaactattacctcgaataaactgcctcag 911
 DB 801 GAGTTTGAAGGCAACTACAGTGGTTTATCGATTACTATCTGGATCAGATGAAGAATGC 860
 QY 912 ctgcgaagcgaagcaagagatgctggtatgcttctgagtaactgatactggtatcctgaagc 971
 DB 861 TCATACGAGAACGCGCAAAAGATTGCTCGATGCTGACGTCCTCATCTTCGGAGC 920
 QY 972 gatggcgaggagcatacgaattacaaatga---ggtaggcgaacgacgaacgaagaagc 1028
 DB 921 ACAGGGCGGAGGCGAAGTTCGTTTGGCGGGCGGCGCAATATCGATACGAGAGGC 980
 QY 1029 cagaatgcaagcgcctgacttctggtggaatccgacacctacaagaagatagctgagtcg 1088
 DB 981 TCGCGTACAAAGCGCCCAAGATCGCTATGGATCCCGCTTACCAGGAAGACAGCTGGATCG 1040
 QY 1089 tcaatggaacagtgaaattcttgccttactgcctcgtatgaagcagtcggtggataagta 1148
 DB 1041 CACATGGTTTCAAGCTACTTGGCCCTTAATTTCCGAAGCTGCAATCTTCGATTCAGAGCTA 1100
 QY 1149 ttaccgggaaacaaagctggttggactgagtagctatggtgcaaaaatgatatctc 1208
 DB 1101 TTATCCGGGTACGAAGCTGGCGATCACAGACTTCAGCTACGCGGAGACATACATCTTC 1160
 QY 1209 cggcggtatcgtatgctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1268
 DB 1161 GGGAGGCATACCTACCGCGGACGCGCTCGGCAATTTTGGAAATATGTCGCTTTATGCGCG 1220
 QY 1269 aaactactggaagttaagagatggtgcaacaactagtagtcccttacaagcttta 1328
 DB 1221 GAATTTACTGGCAG---ACGGAGGACAAATACCGATTTATACGACGCTGCTTCAAGCTGTA 1277
 QY 1329 ccgcaattatgacggaaaaagctctacttctggtgatatcagcgttcatgcgcaaacgtc 1388
 DB 1278 TCGCAACTACGACGCAATAAATCGGGTTTCGGCTCGATCAAAAGTGGAGCGCGCTACGTC 1337
 QY 1389 ggatatgttaatagctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1448
 DB 1338 CGATACGAGGAGAACAGCTCGGTATACGCTTACGCTAACCTGACGAGGAGAAATTCGCAACTCCA 1397
 QY 1449 cctcgtgtcatgaataaaagcagtgacagtgcatgcagcgcacccaatttgatcttccgg 1508
 DB 1398 CCTGATCGTGTGAATAAATAATTTTCGAGGATCCGATCAACGCTACTTTCCAGCTGCTGG 1457
 QY 1509 cgagacgacttaccggttccggtaagtggttgggttgggttgggttgggttgggttgggttgg 1568
 DB 1458 TGATAAAACCTTACACATCCGGGAGATGATGGGCTTCGACCAAAACCGGATCCGACATTAC 1517
 QY 1569 ggaagcagcgcacacacacacacacacacacacacacacacacacacacacacacacacac 1628
 DB 1518 GGAACAGCAGCTATACGAAATATTAACAACAATCAATCAATCAATCAATCAATCAATCAAT 1577
 QY 1629 gacggttctacacatcgttggactgctgctgctgctgctgctgctgctgctgctgctgct 1673
 DB 1578 GTCGGCTTACCACATTTGTTCTGAAAGCGGATAGCACCGACCGGT 1622

RESULT 3

US-08-479-939-28
 ; Sequence 28, Application US/08479939
 ; Patent No. 5685593
 ; GENERAL INFORMATION:
 ; APPLICANT: Woeldike, Helle F.
 ; APPLICANT: Hagen, Frederick
 ; APPLICANT: Hjort, Carsten M.
 ; APPLICANT: Sven, Hastrup

Db 1398 CCTGATCGTCTGAATAAATTTTCGACGATCGATCGATCAACGCTACTTTCCAGCTGCTCGG 1457
QY 1509 cgaagacactacgttcggttaaaagtatgggttttcgacaaaataagctcgcaaatataa 1568
Db 1458 TGATAAACCCTACACATCGCGGAGAGATATGGGCTTTTCGACCAACCGGATCGGACATTAC 1517
QY 1569 ggaagcagcgcccaatcaccaaatcaggcaacggytttacctatatacagtagccgcttt 1628
Db 1518 GGAACAAGCAGCTATAACGATATTAACAAACATCAATTCAGTATACGCTTCCCTCCATT 1577
QY 1629 gacggttatcacatcggtttgactccgagcaatgatacacactgt 1673
Db 1578 GTCGGCTTACCACATTTGTTCTGAAAGCGGATAGCACCAGCGGT 1622

RESULT 4

US-08-483-432-28
; Sequence 28, Application US/08483432
; Patent No. 5763254
; GENERAL INFORMATION:
; APPLICANT: Woelldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; TITLE OF INVENTION: or Hemicellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5763254o No. 5763254disk of No. 5763254th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,432
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE:
; APPLICATION NUMBER: US 07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus/Humicola insolens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1851

US-08-483-432-28

Query Match 15.5%; Score 629.4; DB 1; Length 1854;
Best Local Similarity 63.8%; Pred. No. 1.9e-200;
Matches 986; Conservative 1; Mismatches 552; Indels 6; Gaps 2;

QY 132 tgttaccattaaagtcgatacatccaaggatcgtaagcctattagcccttatatttcagg 191
Db 81 TGTCACTTTCACGATTATATACGCGATCGGAACGTGACGATCAGCCCAATATTTACGG 140
QY 192 tacgaatcagagttggcagcgatgagaactctgactgcgaagacttgggtgcaatcg 251
Db 141 AACCAATCAGGATCTGAGCGGACGGAGAACTGTGTATCCCGAGGTCGGAGCAACCG 200
QY 252 aatgacgggataaactgggaaacaatatgtccaatgcaggaagcagctgtagtcagtc 311
Db 201 GCTGACGGTTTACAACCTGGGAGAAACAACGCATCCAGCGCGGAGGAGGACTGGCTTCATTA 260
QY 312 cagcgatagctattatgcgacaacgcggtatgcacaaaagccgaatgtgaaaaagccagg 371
Db 261 CAGCGATGATTTTCTCTCGCGCAACGGTGGTGTTCAGACACACCGACTCGCAACAGCGGG 320
QY 372 tgcggtggcaacccctgtttcacgatcaatgcgtgaagcagggcacatatcttttagtcac 431
Db 321 GCGGTTGTTACCGCTTTTACGATAAATCTTTGGAGAAATGGAGCTTACTCCATTGTAAC 380
QY 432 actgcccagatggccgggttatgtggccaaaggatggaaacgggaagtgtgcaggaagcga 491
Db 381 GCTCAAAATGCGGGTTATGTCTCCGGGATAAGTCGAGTTTTCGCAAAATGCGCGGTCTCTCCCTTCA 440
QY 492 ggcctccctccgctcggtgggaatgaggtcgtaaacgctaaaaatgcgcggtttcaattgca 551
Db 441 GGCCTCGCTCACCGGTTGGGATAAGTCGAGTTTTCGCAAAATGCGCGGTCTCTCCCTTCA 500
QY 552 gcctgactcgaagacaatcaggtttatgcggaatgcattcgtcaaatcttttagtgaaaaa 611
Db 501 GCCTGATCTGAACGACGACGACAGTGTATATGGATGAAGAAGTTAACTTCTTGCTCAACCG 560
QY 612 gtacggcgttgccttcaaaaaaacggcggtgaaaggatactcgtcgaacatgaacccgc 671
Db 561 GTATGAAACGCTTCAACGCTCAACGGCATCAAAAGCGTATTCGCTGGATAACGAGCGGC 620
QY 672 tctcgtcgcatacgcacgcgcattcgtgaaaggctcggagcgaagagtggt 731
Db 621 GCTGTGCTGAGACGCATCCAGGATTCATCCGGAGAGTTCAGAGCTTACAGCGCAGAACTCGT 680
QY 732 agaccggtcggtaagtattatccaaagcgcgttaagcgggttgacgcggtgcggaatatt 791
Db 681 CGCTAAGAGCATCGACTTGTCAAGGCGGTGAAGAAGCTCGATCCCGATCCCGAATATT 740
QY 792 tgggcccgttctttaaagcttttggcgctataaagatcttcaaacgtcacctgattggaa 851
Db 741 CGGTCTTCCCTTTTACGGTTTCGGCGCATATTTGCTCTCGAGGACGACACCGGATGGCC 800
QY 852 ctctgtaaaagcaactacagctggttcggaactattacctgatcaaatgagcctcag 911
Db 801 GAGTTTCAAGGCAACTACAGCTGGTTTATCGATTACTATCTGGATCAGATGAAGATGC 860
QY 912 ctgcgaagccggaagagattgctggatgctcttcgattgtagtacctggttatcctgaagc 971
Db 861 TCATACGACAGACGCGCAAAAGATTGCTCGATGTGCTGGAGCTCCACTGGTATCCGGAAGC 920
QY 972 gatggcgagggcgcatacgaattacaaatga---ggtaggcaacgaagaaacgaagaagc 1028
Db 921 ACAGGGCGGAGGCGCAGCAATCGTCTTTGGCGGGCGGGCAATATCGATACGCAAGAGC 980
QY 1029 cagaatgcagcgccctcgtactttgggataccgacctacaagaagagatagctgagtcgc 1088
Db 981 TCGGTATCAAGCGCCAAAGATCGGTATGGGATCCGGCTTACCAGGAAGACAGCTGGATCGG 1040
QY 1089 tcaatggaacagtgaaattcttctgcttactgctcgtcgtcgttataaagcagtcggtgataagta 1148

Db 1041 CACATGGTTTTCAAGCTACTTGCCTTAATTCGAAGCTCAATCTTCGATTTCAGACGTA 1100
Qy 1149 ttaccgggaaccagctggttctgactgagtagatagctatggtggcgaataatatttc 1208
Db 1101 TTATCCGGGTACGAAGCTGGCGATCAGAGATTTCAGCTACGGCGGAGACAACTACATTTT 1160
Qy 1209 cggcggtatcgtatggtccggtatggtggtggtggtggtggtggtggtggtggtggtggt 1268
Db 1161 GGGAGGCAATAGTACCGCGGACGGCTCGGCATTTTGGAAATATATGCGCGTTATGCCCG 1220
Qy 1269 aaactactggaagttaagtagtgcgaacaaactagcttagtgcgcttacaagcttta 1328
Db 1221 GAATTACTGGCAG---ACGGAGGACAATACCGATTATACAGGCGCTCTTACAAGCTGTA 1277
Qy 1329 cggcaattatgcaaaaaagcttacttcttctggtggtatgctagcgttcatcgcaaacgtc 1388
Db 1278 TCGCAACTACGACGCGCAATAAATCGGGGTTCGGCTCGCATCAAGGTGGAACGCGCTAGCTC 1337
Qy 1389 ggaattattgtaagctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1448
Db 1338 CGATACGAGAGACAGCTCGGTATACGCTTCGGTAACCTGACGAGGAGAAATCCGCAACTCCA 1397
Qy 1449 cctggttgcataaataaagcattggaacagtgatgctgacggtggtggtggtggtggtggtggt 1508
Db 1398 CTTGATCGTGTGTAATAAATTTTCGACGATCCGATCCGATCAACGCTACTTTTCCAGCTGTCTGG 1457
Qy 1509 cgaacgacttaccgttccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1568
Db 1458 TGATAAACCTACATCCGCGAGAGTATGGGGCTTCGACCAACCGGATCCGACATTAC 1517
Qy 1569 ggaagcagcgcacacacacacacacacacacacacacacacacacacacacacacacacacacac 1628
Db 1518 GGAACAAGCAGCTATAAGCAATATTAAACAACAAATCAATTCAGGTATACGCTTCCTCCATT 1577
Qy 1629 gacggcttatcacatcggttgcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1673
Db 1578 CTCGGCTTACCACATTTGTTCTGAAAGCGGATAGCACCAGAACCGGT 1622

RESULT 5

US-09-136-574A-2

Sequence 2, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing Fabrics Using Truncated

Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6416 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2

Query Match 10.3%; Score 417.4; DB 4; Length 6416;
Best Local Similarity 56.3%; Pred. No. 7.4e-129;
Matches 893; Conservative 1; Mismatches 642; Indels 51; Gaps 4;

Qy 132 ttttaccattaaagtcgatacaccagagatcgtaagcctatttagccttattatttcagg 191
Db 4329 TGTAAATATATCGATCGATACGTCAGGCGGAAAGACAAAGATAAGCCCGTATATTTATGG 4388
Qy 192 tacgaatcaggagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 251
Db 4389 AGCAATCAGGATATCCAGGCTGTTGTTCCACC--CTGCAAGACGACTTGGTGGGAACG 4445
Qy 252 aatgacggatataaactggggaacaaatgttccaatgacgaagcagctggtggtggtggtggt 311
Db 4446 ATTGACGGGTAAATTTGGGAGAACAAATATGTCCTAATGAGGAGAGTACTGCTGATCATTC 4505
Qy 312 cagcgatagctattatgcgaacacgcgcggtatgcgaacacgcgcggtatgcgaacacgcgcggt 371
Db 4506 AACCGATGATATATGTTGTTATATATGTTGTTATGTTGTTATGTTGTTATGTTGTTATGTTGTT 4565
Qy 372 tgcggtggcaacccctgttccagatcaatcgctggaagcaggggcacacatattcttttagtcac 431
Db 4566 AGCTGTTGTAAGCAAAATTTTACGAGCAGTCAATAAAGAAAATGCAATATTCAGCCATCAC 4625
Qy 432 actgacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 491
Db 4626 ATTACAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4685
Qy 492 ggtccttcctgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 551
Db 4686 AGTCCGTCGCCGAGATGGCTGAGGTCAAGTTTAAAAAAGATGGTGACATGTCATTGCA 4745
Qy 552 gctgacgtgaaagacaatcaggtttatcggtggtggtggtggtggtggtggtggtggtggtggtggt 611
Db 4746 GCCTGACGTGAATGATAAATATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 4805
Qy 612 gtacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 671
Db 4806 GTATGTTGATCATCTGCTGCAACGGGAATTAAGGATATATACTTGACACAGGCGCGGA 4865
Qy 672 tctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 731
Db 4866 CTTATGTTTACTACTCCCGCAATTCATCCACAGAGGTAACCTGCAGTGAATTGAT 4925
Qy 732 agaccggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 791
Db 4926 AAATAAATCGGTGGAGCTGGCGAAAGTAATAAAGACACTTGATCCAGATGCAAGAAATTTT 4985
Qy 792 tgggcccgttctttagcgttttggcgcctataaagatcttcaaacctgcacctgattggaa 851
Db 4986 TGAACCTGCATCGTATGTTTGTGGGATATTTAACTTCCAGGATGCACTGACTGGA 5045
Qy 852 ctctgtataaaggcaactacagctggttctgtggtggtggtggtggtggtggtggtggtggtggtggt 911
Db 5046 TCAGGTTAAAGGAAATCACAGATGGTGTGTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5105

Query Match	2.1%	Score 85.2	DB 4	Length 2029
Best Local Similarity	49.1%	Prod. NO. 1.1e-17		
Matches 225	Conservative 0	Mismatches 233	Indels 0	Gaps 0
QY 3591	aacggggatgggtcccgatctgcacagcgagatacaaatgtgaacagcaatcaactgaa	3650		
Db				
	6	AAAGTGGTGTGAAGGTACTGTACAGACACATGAGACAAGTGGCAGCAGAGTTCATTAAG	65	
QY 3651	cccgcatctccaatttttaataaagggtacaaatctccgtaccgattcaacagagtgtgaaaat	3710		
Db				
	66	GCCTGTGGTTTAAGATAGTAGTGAATGAGGACGACGAGCTTGATCTTAGCAGGGTTAAGAT	125	
QY 3711	tcgctactactactacgattccagctgaccgtgagcagacattcaactcgacgactatgcggt	3770		
Db				
	126	AGATACTGTTACACAGTGGATGGTGACAGCCACAGAGTGGCGGTATGTGATGGCGCAC	185	
QY 3771	gctgagctgctcgaaagtgaatggtagctgggtttaaaatgggataaagctgcacacgggtgc	3830		
Db				
	186	GATAGGGCCAAAGCAATGTGCATCTCAATTTGTGAAGCTTAGCAGCGGAGTCAGTGGAGC	245	


```

; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Artificial Sequence: fusion construct
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; FEATURE:
; OTHER INFORMATION: DNA encoding a Cellulose Binding Domain (CBD)-Gaussia luciferase
; US-09-277-716-21

```

Query Match	0.9%	Score 35.8	DB 4	Length 1146
Best Local Similarity	53.1%	Pred. NO. 0.26		
Matches	76	Conservative 0	Mismatches 67	Indels 0
Gaps	0			
QY 3901	attcaaaactcgtattcataaagcagactggtcgaaactataacgaaagtgcagattactcg	3960		
Db 328	attcagaagaataacaaaatcagatcgtcgaaactacacatacaaaatgactattca	387		
QY 3961	tataaaggtacgcagacttcattcgacgattcacacgaagctcgtgtatcacatggc	4020		
Db 388	tttgatgcaagttagttcaacaccagttgtaaatcccaagtattacaggatataggtgga	447		
QY 4021	gtacttgtttgtgggaaccggaacc	4043		
Db 448	gctaaagtctctgtgtacagcacc	470		

```

RESULT 13
US-09-007-005-17
US-09-007-005B
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihe
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
NAME/KEY: misc_feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

Query Match 0.98; Score 34.8; DB 4; Length 289;

[illegible]

```

RESULT      14
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
; US-09-244-796-17

```

[illegible]

Search completed: July 13, 2002, 00:47:34
Job time: 40055 sec

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:

TELEX: 66141 PENNIE
INFORMATION FOR SEO ID NO: 1.

NAME/KEY: cnc

LOCATION: 1..486
-048-164A-1

Query Match 0.88; Score 34.2; DB 1; Length 486;

Best Local Similarity 52.4%; Pred. No. 0.47;

Matches	75;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

.....

Db 403 TTTGATGCAAGTACTTCAACACCAGTTGTAAATCCAAAAGTTACAGGATATATAGGTGGA 462

— — — — —

Db 463 GCTAAAGTACTTGGTACAGCACC 485

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2002, 13:39:35 ; Search time 11269 Seconds
(without alignments)
4857.889 Million cell updates/sec

Title: US-09-784-554B-3
Perfect score: 4056

Sequence: 1 atgaagcggaataatagtag.....ccgaacgcagcagtaataa 4056

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	1.0	1101	12	CNS00KK2
c 2	39.6	1.0	551	10	BI162075
c 3	39.6	1.0	552	10	BI168835
c 4	39.6	1.0	580	10	BI238998
c 5	39.6	1.0	617	10	BI371260
c 6	39.6	1.0	620	10	BI371548
c 7	39.6	1.0	649	10	BI577592
c 8	38.8	1.0	428	10	R26423
c 9	38.4	0.9	331	9	BB121306
c 10	38.4	0.9	588	12	CNS004M4
c 11	38.4	0.9	1085	12	CNS05DG3
c 12	38.2	0.9	453	9	AI297564
c 13	37.4	0.9	490	12	AQ294083
c 14	37.2	0.9	234	9	AV025859
c 15	36.8	0.9	298	9	BB116519
c 16	36.4	0.9	467	12	AQ611399
c 17	36.4	0.9	651	12	C111B9

18	36.2	0.9	449	10	BF602750
19	36.2	0.9	546	9	BE237451
20	36.2	0.9	564	10	BG590175
21	36.2	0.9	671	12	AG123669
22	36.2	0.9	1101	12	CNS000D1
23	36	0.9	424	10	BI706601
24	36	0.9	753	10	BE253470
c 25	35.8	0.9	373	9	AI008625
26	35.8	0.9	387	9	AA800875
27	35.8	0.9	409	10	BI276135
28	35.8	0.9	409	10	BF417754
29	35.8	0.9	423	10	BF411329
30	35.8	0.9	433	10	BF407259
31	35.8	0.9	452	10	BI287601
32	35.8	0.9	458	10	BE459078
33	35.8	0.9	491	10	BF419225
34	35.8	0.9	493	9	AW929704
c 35	35.8	0.9	507	9	AW913846
36	35.8	0.9	508	10	BG590005
37	35.8	0.9	524	9	AI013076
c 38	35.8	0.9	575	9	BE037033
39	35.8	0.9	582	10	BI285977
40	35.8	0.9	589	10	BG589848
41	35.8	0.9	607	9	AI012430
42	35.8	0.9	628	9	AA851940
43	35.8	0.9	660	9	AI177412
44	35.8	0.9	672	12	BH039441
45	35.8	0.9	696	10	BI933811

ALIGNMENTS

CNS00KK2 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR17F22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL077673
AL077673.1 GI:4957249

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

Direct Submission

Submitted (02-JUN-1999)

Genoscope - Centre National de Sequencage;
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"


```

/clone="RE09728"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT      135 a   142 c   144 g   131 t
ORIGIN

```

```

Query Match      1.0%; Score 39.6; DB 10; Length 552;
Best Local Similarity 50.5%; Pred. No. 5.4; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 0;

QY 505 cgggtgaatgaggtcgtaaaacgtaaaacgcccgtttcaattgagcgcctgatctgaaa 564
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 CTGTGGAACGCCAACCTTTCAGCGACAGATTGTTTCATTCTCGTCCAAACCGATTAAAT 146
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 565 gacaatcaggtttatcggtgaattcgtaacttttttagtgaaaaagtagcgggtgct 624
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 TCAAAATCAGATTTCACCTGCTAGAGCTCGATTTCCTTCTGTTGACGTTGAC 86
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 625 tcaacaaaaacggcggtgaaggtactcgctcgacaatgaacccgctctctggtcgcat 684
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 ACACGTGATATCTCGGGGGAACGAACCTGCTTCTAGCGGGAACCGACTCAAGTTCCAC 26
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 685 acgcatccgc 694
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 ACTCAACGGC 16

```

```

RESULT 4
BI238998/c
LOCUS
DEFINITION
  RE35376.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE35376 5 similar to Cyp301a1:
  FBan0008587 'cytochrome P450' located on: 2R 49B2-49B2; : 05/12/2001
  , mRNA sequence.
ACCESSION
  BI238998
VERSION
  BI238998.1 GI:14707503
KEYWORDS
  EST.
SOURCE
  fruit fly.
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 580)
AUTHORS
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
  ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
  ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
  Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
  ,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AE003821: arm:2R [7526095,7816768]
  estimated-cyto:49A12-49D1: 05/12/2001
  Plate: RE.353 row: G column: 4
  High quality sequence stop: 472.
  Location/Qualifiers
  1..580
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"

FEATURES
  source

```

```

/clone="RE35376"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT      142 a   146 c   150 g   141 t
ORIGIN

```

```

Query Match      1.0%; Score 39.6; DB 10; Length 580;
Best Local Similarity 50.5%; Pred. No. 5.6; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 0;

QY 505 cgggtgaatgaggtcgtaaaacgtaaaacgcccgtttcaattgagcgcctgatctgaaa 564
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 CTGTGGAACGCCAACCTTTCAGCGACAGATTGTTTCATTCTCGTCCAAACCGATTAAAT 146
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 565 gacaatcaggtttatcggtgaattcgtaacttttttagtgaaaaagtagcgggtgct 624
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 TCAAAATCAGATTTCACCTGCTAGAGCTCGATTTCCTTCTGTTGACGTTGAC 86
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 625 tcaacaaaaacggcggtgaaggtactcgctcgacaatgaacccgctctctggtcgcat 684
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 ACACGTGATATCTCGGGGGAACGAACCTGCTTCTAGCGGGAACCGACTCAAGTTCCAC 26
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 685 acgcatccgc 694
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 ACTCAACGGC 16

```

```

RESULT 5
BI371260/c
LOCUS
DEFINITION
  RE58119.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE58119 5 similar to Cyp301a1:
  FBan0008587 GO:[] located on: 2R 49B2-49B2; : 05/16/2001, mRNA
  sequence.
ACCESSION
  BI371260
VERSION
  BI371260.1 GI:15067288
KEYWORDS
  EST.
SOURCE
  fruit fly.
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 617)
AUTHORS
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
  ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
  ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
  Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
  ,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AE003821: arm:2R [7526095,7816768]
  estimated-cyto:49A12-49D1: 05/16/2001
  Plate: RE.581 row: B column: 7
  High quality sequence stop: 473.
  Location/Qualifiers
  1..617
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"

FEATURES
  source

```



```

/clone="RE71294"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/notes="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT      158 a 168 c 167 g 156 t
ORIGIN

```

```

Query Match      1.0%; Score 39.6; DB 10; Length 649;
Best Local Similarity 50.5%; Pred. No. 5.9;
Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 505 cgggtgaatgaggtcgtaaacgctaaataatgcgcgtttcaattgcagctgacgtgaa 564
      |||||
Db 208 CTGTGGAAAGCCCAAGCTTCAGCGACAGATTGTTCACTCTGTCACCAACCGATTAAAT 149

QY 565 gacaaatcaggtttatgcggatgaattgcacacttttttagtgaataagtcacgcgttgct 624
      |||||
Db 148 TCAATCAGATTTCACATGGTTCGTAAGCCTCGATTTCACATTTCTTCTGTGACGTTGAC 89

QY 625 tcaacaaacagggcgtgaaaggatctcgtcgacaaatgaacccgcgtctctctggtcgcat 684
      |||||
Db 88 ACACGTGATATCTGCGGGGAACGAACAACTGGCTTCTAGCGGAACCCGCACTCAAGTTCCAC 29

QY 685 acgcatccgc 694
      |||||
Db 28 ACTCAACGCC 19

```

```

RESULT 8
R26423/c
LOCUS      YH47c11.rl Soares placenta NB2HP Homo sapiens cDNA clone
DEFINITION IMAGE:132884 3', mRNA sequence.
ACCESSION R26423
VERSION    R26423.1 GI:782558
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 428)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
           M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
           Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
           R., Williamson,A., Wohldmann,P. and Wilson,R.
           The Washo-Merck EST Project
           Unpublished (1995)
           Contact: Wilson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Insert Size: 668
           High quality sequence stops: 144
           Source: IMAGE Consortium, LLNL
           This clone is available royalty-free through LLNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Insert Length: 668 Std Error: 0.00
           Seq primer: -21m13
           High quality sequence stop: 144.
           Location/Qualifiers
           1..428
           /organism="Homo sapiens"
           /db_xref="GDB:538536"
           /db_xref="taxon:9606"

```

```

/clone="IMAGE:132884"
/clone_lib="Soares placenta NB2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AAGTGGAGAAATTCGCGCGCAGCAATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT      109 a 80 c 82 g 144 t 13 others
ORIGIN

Query Match      1.0%; Score 38.8; DB 10; Length 428;
Best Local Similarity 50.0%; Pred. No. 8;
Matches 88; Conservative 1; Mismatches 87; Indels 0; Gaps 0;

QY 1435 tacaaagaactgcacactgttgcacgaataaaacatgcagcagtcgacgtgcacgcccaa 1494
      |||||
Db 314 TNGACAAATNGCCACCAATTTTAAACCAAAAGGAAGGAAGTGCATATAGCACCTAA 255

QY 1495 ttgatcttttcggcagcagcagcttacggttcggttaaagtatgggttcgacaaaaa 1554
      |||||
Db 254 AGAGGTGTTTCCATCAAAATGCTTAAAGGATTTTAAATAATGAGGACGAGAACAAAT 195

QY 1555 agctcgcaaataaagcagcagcgcacaaatcaccaaatcaggcgaacgcggtttac 1610
      |||||
Db 194 CACTGTGATGATGAACCTGTTGTAATGAAAAAGTCCCGAGCAAGTGTGTAC 139

RESULT 9
R26423/c
LOCUS      BB121306
DEFINITION BB121306 RIKEN full-length enriched, adult male urinary bladder Mus
           musculus cDNA clone 9530081106 3' similar to AF102578 Mus musculus
           intracellular chloride channel protein (Cllic) mRNA, nuclear gene
           encoding mitochondrial protein, mRNA sequence.
ACCESSION  BB121306
VERSION     BB121306.1 GI:8773874
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 331)
AUTHORS    Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
           P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
           Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
           Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
           Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
           Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
           Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y.,
           Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
           Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toyai
           T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
           Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
           M., Muramatsu,M. and Hayashizaki,Y.
           RIKEN Mouse ESTs (Konno.H., et al.)
           Unpublished (2000)
           Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan,
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gs.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/

```

SOURCE	ORGANISM	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
fruit fly.	Drosophila melanogaster		
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 588)			
Genoscope.			
Direct Submission			
Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage :		
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).			
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila/bac.htm .			

melanogaster genome using these BACs. For further information please see <http://www.flylib.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pletter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

[illegible][illegible][illegible]

```

DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
053D02 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL332364.1 GI:8226122
VERSION AL332364
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1085)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1085)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1085)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1.1085
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="053D02"
/clone_lib="A"
/note="Genoscope sequence ID : COAA053DB01A1-end : T3"
BASE COUNT 299 a 260 c 302 g 210 t 14 others
ORIGIN

Query Match 0.9%; Score 38.4; DB 12; Length 1085;
Best Local Similarity 54.2%; Pred. No. 18;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1969 gaagtgacagcctgaaggaccgttgtgaatccagtgagcgcgtcttcctcggtact 2028
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 GAGGACTGCAAACTGAGACGGAGCCGCGGAGAGCCGGTGGATCGGGTTCTCGGGTAC 604

QY 2029 ggtatgtaactaattccacaatgcagggtattccctgcagatgacatccaggctccc 2088
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 603 CGCATCTACCGACGCGCTGGGCTCTCTCGGCTCCCGAGCTGCTCCGCTCGCTCGC 544

QY 2089 acggcaggctgtacaatttaca 2112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 ACGGCACCTCGGTTAATGGGCACA 520

RESULT 12
AI297564/c 653 bp mRNA linear EST 19-APR-2001
LOCUS LP11891.5prime LP Drosophila melanogaster larval-early pupal pot2
DEFINITION Drosophila melanogaster cDNA clone LP11891 5prime, mRNA sequence.
ACCESSION AI297564
VERSION AI297564.1 GI:3946971
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 653)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 118 row: H column: 7
High quality sequence stop: 455.
Location/Qualifiers
1..653
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP11891"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pot2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."
BASE COUNT 158 a 165 c 170 g 160 t
ORIGIN

Query Match 0.9%; Score 38.2; DB 9; Length 653;
Best Local Similarity 50.3%; Pred. No. 15;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 505 cggtggaatgaggtcgctaaacgctaaataatgcgcccgtttcaattgcagcctgactgaaa 564
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 187 CTGTGGNACCCCAAGCTTTCAGCGACAGATTGTTCTACTCTGCCAACCGATTAAAT 128

QY 565 gacaatcaggtttatcgagatgaattcgtaacatttttagtgaagaagacgcgctgct 624
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 127 TCAAAATCAGATTTCACCTGCTGCGTAAGCCTCGATTTCCTCTGTTGACGTGAC 68

QY 625 tcaacaaaaacggcgtgaagaagatactcgctgacataacccgctctctgctgcat 684
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 ACATGTATATCTGCGGGGAAACGAAACTGGCTTCTAGCGGGAACCGCACTCAAGTTCAC 8

QY 685 acgcatac 691
|||||
Db 7 ACTCAAC 1

RESULT 13
AQ294083 490 bp DNA linear GSS 15-DEC-1998
LOCUS HS_2191_B2_H06_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-2191 Col-12 Row-P, DNA sequence.
ACCESSION AQ294083
VERSION AQ294083.1 GI:4011666
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 490)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
TITLE
JOURNAL

```

1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:47:40 ; Search time 85.28 Seconds
(without alignments)
1758.321 Million cell updates/sec

Title: US-09-784-554B-4

Perfect score: 7165

Sequence: 1 MKAKNSSIWSKRSKWLPPV.....HTKATLYHNGVLWGVTETAN 1350

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_032802.*
- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7165	100.0	1350	22	AAG63963 Amino acid sequenc
2	6413	89.5	1352	22	AAG63962 Amino acid sequenc
3	3161	44.1	695	22	AAG63964 Amino acid sequenc
4	1816.5	25.4	700	12	RAA13227 Novel endoglucanas
5	1795	25.1	617	12	RAA15241 B.lautus endol cor
6	1524	21.3	1751	20	AA113493 Truncated cellulosa
7	1266	17.7	496	21	AA154135 Amino acid sequenc
8	1263.5	17.6	815	21	AA154134 Amino acid sequenc
9	1105.5	15.4	488	21	AA154128 Amino acid sequenc
10	927.5	12.9	903	21	AA154137 Amino acid sequenc
11	550.5	7.7	551	18	AA118790 Corrected Bacillus

12	446	6.2	782	12	AA15625 Cellulase AE-1. A
13	423.5	5.9	1426	20	AA13492 Truncated cellulasa
14	375	5.2	616	20	AA13494 Truncated cellulasa
15	352.5	4.9	499	14	AA12122 NK-1 cellulase. B
16	339.5	4.7	531	18	AA15238 Scaffoldin protein
17	339.5	4.7	1853	19	AA13108 C. thermocellum ce
18	336.5	4.7	493	20	AA12850 Pectate lyase-link
19	336.5	4.7	493	20	AA13218 Pectate lyase CBD
20	334.5	4.7	476	21	AA154123 A mannanase-linker
21	320	4.5	531	16	AA101503 60 kD endoglucanas
22	311	4.3	532	12	AA13229 Endoglucanase enco
23	303	4.2	167	17	AA13229 Cellulose binding
24	303	4.2	537	18	AA13229 Bankia gouldi endo
25	303	4.2	537	19	AA13229 Bankia gouldi endo
26	283	3.9	362	21	AA14497 Bacillus subtilisi
27	276	3.9	336	21	AA14497 B. subtilis strain
28	271.5	3.8	1475	11	AA108221 Recombinant alpha
29	264.5	3.7	586	21	AA154126 Amino acid sequenc
30	263	3.7	361	21	AA154136 Amino acid sequenc
31	260	3.6	487	10	AA191007 Beta-mannase of al
32	254.5	3.6	341	22	AA105747 Clostridium cellu
33	245	3.4	419	22	AA105747 Mannanase. Pseudo
34	237.5	3.3	428	22	AA105748 Clostridium cellu
35	232	3.2	328	22	AA105749 Chimeric S peptide
36	229	3.2	162	20	AA105777 C. cellulovorans C
37	225	3.1	1010	22	AA105777 Mannanase. Cellul
38	224.5	3.1	470	22	AA105777 Mannanase protein
39	224	3.1	162	15	AA105777 Cellulose binding
40	224	3.1	163	22	AA105777 Clostridium cellu
41	223	3.1	154	20	AA105777 C. cellulovorans C
42	223	3.1	156	20	AA105777 C. cellulovorans C
43	223	3.1	382	20	AA139552 Gaussia luciferase
44	218	3.0	190	22	AA105746 Clostridium cellu
45	212.5	3.0	1291	19	AA105912 Amino acid sequenc

ALIGNMENTS

RESULT 1

AAG63963

ID AAG63963 standard; Protein: 1350 AA.

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

XX AC AAG63963;

PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries -

PS Disclosure; Page 86-89; 97pp; English.

XX
 CC The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rattling of fibers e.g. hemp, flax, and
 CC linen.

XX Sequence 1350 AA;

Query Match 100.0%; Score 7165; DB 22; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKAKNSSISWKRSLPVMVMACTIIVGGALPTPTVVHGTAKTIVTISKDRKPISP 60
 DB 1 mkaknssiswkrslpvmvmaactiivggaltptvvhgtaktvtikvtskdrkpisp 60

QY 61 YIYGTNQLAGDENLTARRLGGNRMWGYNNWNNMNSAGSDMSYLCNAGLTAKEC 120
 DB 61 yiygtngelagdenltarrlgnrmtgynwnmnsagdwmsdmsylcdnagltkaec 120

QY 121 EKPAGVATSFHQSLKQGTYSLVTLPMAGYVAKDNGSVQSEKAPSPARNEVYVNAKNAP 180
 DB 121 ekpgavatsfhdqslkqgtyslvtlpmagyyvakdngsvqsekapsarnevynaknap 180

QY 181 FOLQPDLDKNQVYADEFVNFVLVKKYGVASTKTGVKGYSLDNEPALWHSHTPHRIHGEKVG 240
 DB 181 fqlqpdldknqvayadevfnflvkkgyvastktgvkgySLDNEPALWHSHTPHRIHGEKVG 240

QY 241 KELVDRSVLSLSKAAKAVDAGAEIPEGVLYGFGAYKDLQAPDWNKNGYSEWFDYLDQ 300
 DB 241 kelvdsvlsLSKAAKAVDAGAEIPEGVLYGFGAYKDLQAPDWNKNGYSEWFDYLDQ 300

QY 301 MRLSSQAEGRLLDVFVHVWYPEAMGGGIRITNEVGNDETAKARMOAPRTLWDPITYKEDS 360
 DB 301 mrlssqaeqrlldvfvhvwyPEAMGGGIRITNEVGNDETAKARMOAPRTLWDPITYKEDS 360

QY 361 WIAQWSEFLPLRLKQSDVKYYPGKTLALTEYSYGGENDISGGIAMDVLGILGKNDV 420
 DB 361 wiaqwnsefllprrlkqsdvkyypgkTLALTEYSYGGENDISGGIAMDVLGILGKNDV 420

QY 421 YMANVWKLKDGANNVSAAYKLYRNYDGKSTFGDISVHAQTSIDVSNVSHVASTDASYK 480
 DB 421 ymanvwklkdgannvsaayklyrnydgkSTFGDISVHAQTSIDVSNVSHVASTDASYK 480

QY 481 ELHLVVMNKMDSAFDAQFDLSGETTYTGVKGMVGFQDKNSQIKAEAPITQISGNRRITYV 540
 DB 481 elhlvvmnkmDSAFDAQFDLSGETTYTGVKGMVGFQDKNSQIKAEAPITQISGNRRITYV 540

QY 541 PPLTAYHIVLTAGNDTPVENPESFALRAEAGDGKSIYLDASSGVGVSVORATYENGPPFA 600
 DB 541 ppltayhivltAGNDTPVENPESFALRAEAGDGKSIYLDASSGVGVSVORATYENGPPFA 600

QY 601 AVASNLVETSTDTNTVNTNGSYYYKTKTCTGSASNLVKAIPRAPVDPGRYEAEEDGT 660
 DB 601 avasnlvetstDNTVNTNGSYYYKTKTCTGSASNLVKAIPRAPVDPGRYEAEEDGT 660

QY 661 LKGTWVSSGTGSGTGYVNFHFNAGDSLTMITQAPTAGLNTIGYRSPHDDKRTNFSL 720
 DB 661 lktvwssgtgsgtgyvntfnfnagdslmtitqaptaglnltigyrsphddkrtnfsl 720

DB 661 lktvwssgtgsgtgyvntfnfnagdslmtitqaptaglnltigyrsphddkrtnfsl 720
 QY 721 NGKASGELVLWKTADFKETSGGKVLNAGANTIGFTGWCYDIDVVKLEPAADRPHAV 780
 DB 721 ngkasgelvlwktadfketsggkvlNAGANTIGFTGWCYDIDVVKLEPAADRPHAV 780
 QY 781 TKTLPNPATVEAKALMNYLVDQYGNMISGGQEDMPEIDWLQANVGKPAIAALDLIDYS 840
 DB 781 tktlpnpatveakalmnylvdqygnmISGGQEDMPEIDWLQANVGKPAIAALDLIDYS 840
 QY 841 PSRAEHGLSTETEKAEIWDKQGGIVTFPAWHNAPKGLIDTQCKEWMRGFYADSTTFDIE 900
 DB 841 psraehglstetekaEiwdkqggivtfPAWHNAPKGLIDTQCKEWMRGFYADSTTFDIE 900
 QY 901 YAMNHPSESDYKLLIRIDIVIAQOLKLODAKVPVLPFRPLHEAEGKFWMGAKGPPPVKK 960
 DB 901 yamnhpseedyklliridivIAQOLKLODAKVPVLPFRPLHEAEGKFWMGAKGPPPVKK 960
 QY 961 LYILMHDRLTNVHKLNNLIWVWNSVAPDWPGEYVDILSFDSYPOAGDYSPOIAKYEDL 1020
 DB 961 lyilmhdrltnvhklNNLIWVWNSVAPDWPGEYVDILSFDSYPOAGDYSPOIAKYEDL 1020
 QY 1021 VTILGDKKLVCHEKERTYPPGSDGCVSSPLELVRYWVGFLEROTKOSLEHLKKVYNH 1080
 DB 1021 vtilgdkklvcheKERTYPPGSDGCVSSPLELVRYWVGFLEROTKOSLEHLKKVYNH 1080
 QY 1081 PNVTILEKLPNTLKYTIGITEQSPVPGSFTLNAAGETAKVKLSWTASANAASYEVRKSTVE 1140
 DB 1081 pnvtileklpntlkytIGITEQSPVPGSFTLNAAGETAKVKLSWTASANAASYEVRKSTVE 1140
 QY 1141 NGAFATVASDVGSSYTDPAVTADTTYVYVVAKNADAGQTVSNTASAAPKADTQPTTGL 1200
 DB 1141 ngafatvasdvGSSYTDPAVTADTTYVYVVAKNADAGQTVSNTASAAPKADTQPTTGL 1200
 QY 1201 VLQYRTADTNVNDNLNPHFQILNKGTISVPINELKIRYYTIDGREGTFCNDYAVLSC 1260
 DB 1201 vlqyrtadtnvndnlNPHFQILNKGTISVPINELKIRYYTIDGREGTFCNDYAVLSC 1260
 QY 1261 SKLNGRLVKNMDAAATGADYYLEVSNDSAGVLASGSGTGIGIOTRIHKADWSYNESDDYS 1320
 DB 1261 sklngrlvknmdAAATGADYYLEVSNDSAGVLASGSGTGIGIOTRIHKADWSYNESDDYS 1320
 QY 1321 YKGTQTSFDDHTKATLYHNGVLWGTETAN 1350
 DB 1321 ykgtqtsfddhtkatlyHNGVLWGTETAN 1350

RESULT 2
 AAG63962
 ID AAG63962 standard; Protein; 1352 AA.
 XX AAG63962;
 AC
 XX 29-OCT-2001 (first entry)
 DE Amino acid sequence of xyloglucanase enzyme.
 KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
 OS cellulosic fiber; textile scouring.
 XX Paenibacillus polymyxa.
 PN WO200162903-A1.
 XX 30-AUG-2001.
 PF 21-FEB-2001; 2001WO-DK00116.
 PR 24-FEB-2000; 2000DK-0000291.
 XX (NOVO) NOVOZYMES AS.
 PA Schnorr K, Jorgensen PL, Schuelein M;
 PI

XX WPT: 2001-522819/57.
 DR N-PSDB: AAH/5059.
 XX
 PT New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries
 XX
 PS Example 3: Page 81-85; 97pp; English.
 XX
 CC The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees Celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rating of fibers e.g. hemp, jute, flax, and
 CC linen.
 XX
 SQ Sequence 1352 AA;

Query Match 89.5%; Score 6413; DB 22; Length 1352;
 Best Local Similarity 88.9%; Pred. No. 0;
 Matches 1212; Conservative 48; Mismatches 79; Indels 24; Gaps 7;

QY 1 MKAKNSSSTWKRSLWLPVVMACTIIVGGLPTPTVHVGOTAKVTIKVDFSKDKPLSP 60
 DB 1 mrakussnllfrskwlpvvmactmivggaipapavvhggtaktitkvdtkdrkplsp 60
 QY 61 YTYGNOELAGDENITARRLAGNRMGTGYNNWNNMNSAGSDWMSQSDSYLCNAGLTKAEC 120
 DB 61 ytygnqdlagenmaarrlgnrmgtgynwenmsnagsdwqssdnylcsngltqae 120
 QY 121 EKPQAVATSFHDQSLKQGTYSVLTPMAGYVAKDNGSVQSEKAPSRNNEVYNAP-180
 DB 121 ekpgavttsfhdqslkgtysvltpmagyvakdngsvqsekapswrnvqvnaknap 180
 QY 181 FOQLPDLKDNQVYADFEYVFLVKKYGVASTKTGVKGYSLDNEPALWSTHPRHGEKVA 240
 DB 181 fqlpdlndnrivyvdefvhlvnykystastkagvkgyaldnepalwsthrhgekvga 240
 QY 241 KELVDRSVLSKAAKAVDAGAEIFGPVLYGFAYKDLQATPDWNSVKNGYSNFVDYLLDQ 300
 DB 241 kelvdrsvlskavkaidagaeifgvpvlygfaykdlqatpdwdsvkngyswfvdyldq 300
 QY 301 MRLSSQAEGRLLDVFVHWYPEAMGGIRITNEVGNDETAKARMOAPRTLWDPTKEDS 360
 DB 301 mrlssqvegkrlldvfvhwypeamggiritnevgndetkarmqaprtlwdptkyeds 360
 QY 361 WIAQNSEPLPLPRLKQSVDRKYYPGTKIALTEYSYGGENDISGGIAMADVLGILGNKV 420
 DB 361 wiaqnsefipilprlkqsvdkyypgtklamteysyggendisggiamdvlgilgkndv 420
 QY 421 YMANVWKLKDGANNVSAAYKILRYNDGKSTFGDISVHAQTSIDIVNSSVIASVTDASYK 480
 DB 421 ymanywklldgvnnvsaaykilyrnydgnkstfgdtsvhaqtsidivnsvhasvtnasdk 480
 QY 481 ELHLVVMNKSMDAPDAOPDLSEGTGYGKVGWFGDKSSQIKAEAPITQISGNRFTVT 540
 DB 481 elhlvmmksmdaodaopdlsgaktyisgkvwgrdkussqikeaapitqisgnrftvt 540
 QY 541 PPLTAYHIVLTAGNDT-PVENPESFALRAEAGDGKSYLV--DASSGVVGVSYQVATYENG 597
 DB 541 ppltayhivltgndtspegpesfklkaeagdgk-vhlswdassgvvgyvqratdeng 599
 QY 598 PFAAVASNLVETSYDTNTVNTGTSYYKITAKTKTGTGSASNVLKAVPRAPVDGPDYBAE 657
 DB 598 pfaavasnlvetstntvntgtsyykitaaktkgtgsasnlkavprapvdgpdrybae 657

DB 600 pftavasnlstetsydtntvntgtsyyykvtaktknksgssesnlkavpknmpvngparyeae 659
 QY 658 DGTLCFTVVESSGTGFGSGTYVTNFHNGADSLTWTQAPTAGLYNLITIGYRSPHDDKRTN 717
 DB 660 egtlktivessgtgysgagvtnfnhnpqdsitmtiqaptaglynlitigyrspdhdkrtcn 719
 QY 718 FSLNGKASGELVLWKTADFETSGGKVLNAGANTIGFTGWCWYDIDYVVKLEPAADRRP 777
 DB 720 fslngkafgellllktadfketsggkvlalnagantisfetsgwgwydidyvrlepaadrrp 779
 QY 778 HAVTKLIPNPATVEAKALMNYLDQYGNMLSGQEDMPEIDWLQANQKPKPAIALDLI 837
 DB 780 havtkltpnatveakalmnyldvdygknlmgsgqeineidwlganvgkpaiaaldli 839
 QY 838 DYSPSRAEHLSTETFEKALEWDKQGGIVTFAHWNAPKGLIDTQCKEHWRFYADSTTF 897
 DB 840 dyspsraehglslsteaekalwdkggivtfawhwnapkgldtqgkewrgfyadsttf 899
 QY 898 DIEYAMNHPESEDYKLLIRDIDVIAGQLKLODAKVPVLFPRPLHEAEGKWFWMGAKGPEP 957
 DB 900 dieyamnhpesedykllirididviagqlklldakvpvlfprplheaeagkfwgagkpep 959
 QY 958 VKKLYTLMDHDLRLNVHKLNNLIWVNSVAPDWPYDGEYVDIUSFSDYPOAGDYSPQIAKY 1017
 DB 960 vkklyilmhdrltnvkhlnnliwvnsvapdwpydgyeyvdiusfsdypqagdyspqisky 1019
 QY 1018 EDLYTLGKDKKLYCHERKRTYPGSGDEGVSSPLELVR-----YMWV-GFLERRQTK 1068
 DB 1020 edlvtlkgdkklychervkrtypgsgdegvssplelvr-----dmkayqahswfiatwygdfvrdgkq 1069
 QY 1069 QSLHKLKVVYNHNPVITLEKLPNLKTYGITEQPSVPGSFTLNAAGETAKVKLSWTASAN 1128
 DB 1070 nslhklkvynhnpvitdelplnlktygiteqpsvpgsftlnaagetakvslwtasan 1129
 QY 1129 AASYEYKRVSTVENGAFTVASDYGSSYDTTAVTADTTTYTYVQVANKDAGQTVSNTASAA 1188
 DB 1130 aksyevkrstengafatvasdygssydtavtadttytyvqvankdagqtlntasam 1189
 QY 1189 PKADTQOPTGLVLOVYRTADTNVNDNHLNPHFOILNKGITISVPINELKIRYTYTDGDR 1248
 DB 1190 pkadtqoptglvllyrtadtkvndhlnphfpqilvknkgitispinelkirytytdgdr 1249
 QY 1249 QTENCYDVAVLSCKSLKGLVKMDKAATGADYLYLEVSNFSDAGVLSAGSGSTGGTIRIHA 1308
 DB 1250 qtfnocyatvlscslngklvkmekaatagadylylevsnfsgdagvlsagsgtgdqtrihkt 1309
 QY 1309 DWSYNESDDYSYKGTQTSFDDHTKATLYHNGVLVWGTE-TAN 1350
 DB 1310 dwsynesddysykgtsfadhpkvtlyhngvlvwgteptan 1352

RESULT 3
 AAG63964
 ID AAG63964 standard; Protein; 695 AA.
 XX
 AC AAG63964;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of xyloglucanase enzyme.
 XX
 KW xyloglucanase; family 44; glycosyl hydrolase; detergent;
 KW cellulosic fiber; textile scouring.
 XX
 OS Paenibacillus pabuli.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 239..274
 FT /note= "36 unspecified residues encoded by NNN"
 FT Misc-difference 279
 FT /note= "unspecified residue encoded by NTG"
 XX
 PN WO200162903-A1.

XX 30-AUG-2001.
 XX 21-FEB-2001; 2001WO-DK00116.
 XX 24-FEB-2000; 2000DK-0000291.
 XX (NOVO) NOVOZYMES AS.
 XX Schnorr K, Jorgensen PL, Schulein M;
 XX WPI; 2001-522819/57.
 XX N-PSDB; AAH75061.
 XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PT processing industries
 XX Disclosure; Page 90-92; 97pp; English.
 XX The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10, at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees celsius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
 CC linen.
 XX Sequence 695 AA;

Query Match 44.1%; Score 3161; DB 22; Length 695;
 Best Local Similarity 86.1%; Pred. No. 3.4e-183;
 Matches 608; Conservative 18; Mismatches 56; Indels 24; Gaps 4;

QY 1 MKANSSSIWKRKWLPPVVMACIIIVGGALPTPTVHGHTAKTIVTIKVDTSKDRKPISP 60
 DB 1 mkaknsnlskrskwlpvmmactliivggalpaptvvhgaktvtikvdtksdrkpisp 60
 QY 61 YIYCTNQLAGDENLTARRLCGRMTGYNNENMSNAGSDWQSSDYLCDNAGLTAKAC 120
 DB 61 yiygtngqlagdenlaarrrlggrmtgynwenmnsnagsdwqssdnlcnngltkaec 120
 QY 121 EKPAVATSFHDQSLKQCTYSLVTLPLMAGYVAKDNGSVQSEKAPSAKRNVEVNNKAP 180
 DB 121 ekpavttsfhdqslklgayslvtlplmagyvkdngsvqseqapsarwnqvnnaknap 180
 QY 191 FOLQPDLDKNOVYADEFVNFVLKRYGVASTKTYGKYSLDNEPALWSTHTPHRIHGEKVGA 240
 DB 181 fqlqpdldndqvadevfnflvkkygaastkagkygvaldnepalwshthphrihgekvxx 240
 QY 241 KELVDRSVLSLKAADAGAEIPGPV-----LYGFCAYKDLQOTAPDWNWSVKNYS 291
 DB 241 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxkhlxig-----Lsvkgnys 289
 QY 292 WFDVYILDQMLRSQAEGKRLLDVDFVHWYPEAMGGGIRITNEVGNDETCKKARMQAPRTL 351
 DB 290 wfdvdyldqmrlnsqaegkrlldvdfvhwypeamgggirtnevgndetckkarmqaprtl 349
 QY 352 WDPTYKEDSWIAQNSSEFLPLRLKQSDVKYYPGKTLALTEYSYGGENDISGCIAMADV 411
 DB 350 wdptykedswiaqnsaflplrlkqsdvkyypgkltalteysyggendisgiamtdv 409
 QY 412 LGIILGKNDVYMANVWKLKDCANNVVSAAKLYIRNYDCKSTFGDISVHAQTSDIVNSSVH 471
 DB 410 lgilgkndvymanywklkdgannvvsaaaklyirnydgknatfgdisvnaqtsdivnssvh 469

QY 472 ASVTDASYKELHLVVMNKSMDSAFDAQFDLSGETTYSGKVGWFDKNSSOIKEAAPITOI 531
 DB 470 asvtdasykelhlivvmksmdsafdaqfdlsgettyssgkiwfdknssgikavapitqi 529
 QY 532 SGNRFTVTPPLTAYHIVLTAGNDT---PVENPEFALRAEAGDGK--SIYLDASSGVVGY 587
 DB 530 sgnrftytppltayhivltadndtpvppvedpesfclraeagdgkvdlsdassgvvgy 589
 QY 588 SVORATYENGPFRAVASNLVETSYTDNVTNCTSYYYKITAKTKTGTSTASNVLKAVPRAP 647
 DB 590 svqratyengpfraavasnlvetsytdntvngtsyyykitaktkagtsesnvikavprtp 649
 QY 648 VDGPDRYEADGTLKGTVVESSTGTFSGTGYVTFNHNAGDSLMTI 693
 DB 650 vdpdryeaeagtlkgtivessgtgsgtyvtnfnhagdsldmtl 695
 RESULT 4
 ID AARI3227
 XX AARI3227 standard; Protein; 700 AA.
 AC AARI3227;
 XX 14-OCT-1991 (first entry)
 DT Novel endoglucanase.
 DE Cellulase activity; detergent.
 XX Bacillus spp. NCIMB 40250.
 OS Bacillus spp. NCIMB 40250.
 FH Key Location/Qualifiers
 FT Peptide 1..31
 FT Protein /label= signal sequence
 FT 32..700
 FT /label= mature endoglucanase
 FT Cleavage-site 31..32
 XX WO9110732-A.
 PN 25-JUL-1991.
 PD 18-JAN-1991; 91WO-DK00013.
 XX 19-JAN-1990; 90DK-0000164.
 PR (NOVO) NOVO NORDISK A/S.
 PA Jorgensen PL, Schulein M, Hansen C;
 PI WPI; 1991-238020/32.
 DR N-PSDB; AAQ13001.
 XX Enzyme exhibiting cellulase activity from Bacillus sp. - is an
 PT endo-glucanase, esp. useful for harshness redn. of cotton-contg.
 PT fabrics.
 PS Claim 1; Page 80; 96pp; English.
 XX The enzyme is produced by a strain of Bacillus spp. NCIMB 40250
 CC and exhibits an endoglucanase activity of at least 10 (pref. at
 CC least 25) carboxymethyl cellulose (CMC) endoase units per mg total
 CC protein under alkaline conditions. It is especially useful as a
 CC cellulolytic agent and has been found to be more stable during
 CC washing (60 mins. at 40 deg.) in the presence of conventional
 CC detergents than a commercial cellulase preparation. It may also
 CC show increased storage stability in liq. detergents contg.
 CC proteases. The sequence was deduced from the DNA (AAQ13001).
 CC See also AARI3228 and AARI3229.
 XX Sequence 700 AA;

Db 388 ggiatadalgifgkyvyaanywqtedn-tdytsaayklyrnydgnksfgfsikvdaats 446
 QY 464 DIYNSVSHASVTDASYKELHLVYMNKMSDSDAFDAQDLISGETTYGSGKVMGFDKNSOIK 523
 Db 447 dtenssvyavstdeensehlhivlnknfdpinaatfqlsgdktytsgrwgfddqgsdit 506
 QY 524 EAAPITQISGNREFTYVPLTAHYHVLVTRAGNDTPVENPESFALRAEADGKSIYLDASSG 583
 Db 507 eqaaifninnnftylpllsayhivlkadstepwise----- 544
 QY 584 VVGYSVQRATYENGPFPAASVNLVSTYDNTVNTGTSYYKTKTAKTGTGTSASNLKAV 643
 Db 545 -----ipsstsspvncqpts-----tstststssppvq-- 574
 QY 644 PRAPVDGPPRYBAEDGTLKGTVVESSGTGFSG 675
 Db 575 ptp-----sgctaerwaqcgngwsg 596

RESULT 6

AAY13493
 ID AAY13493 standard; Protein: 1751 AA.

AC AAY13493;

DT 30-JUL-1999 (first entry)

DE Truncated cellulases comprising amino acid sequence.

KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;
 Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;
 cotton-containing fabric; stonewashing.

OS Unidentified.

PN EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-0810919.

XX 19-SEP-1997; 97US-0932571.

XX (CLRN) CLARIANT FINANCE BVI LTD.

XX Anderson P, Bergquist PL, Daniels RM, Farrington GK;
 XX Gibbs MD, Morgan H, Williams DP;

XX WPT; 1999-315403/27.
 XX N-PSDB; AAX55662.

XX New truncated cellulase proteins, useful in detergents and for
 XX producing 'stonewashed' denim

XX Claim 7; Page 37-41; 65pp; English.

CC The invention relates to a recombinant cellulase active protein free of
 CC proteinases of native thermophilic and alkaliphilic origin, comprising
 CC the truncated sequences Cel B5, Cel B4/5, Cel E1, Cel E1/2, Cel 1/2/3,
 CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-
 CC length sequences, or functional equivalents. Cel B5 extends from amino
 CC acid K635 to N1424 or K1425 or N1426, and Cel B4/5 extends from amino
 CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends
 CC from amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel
 CC E1/2/3 extends from Y39 to G812, Cel B6 extends from amino acid V1233 to
 CC K1751 and the stability region extends from amino acid E482 to G635 in
 CC the sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new
 CC enzymes are useful in laundry detergent compositions to prevent or
 CC remove staining, backstaining or graying, for use on cellulosic
 CC materials including cotton-containing fabrics. They are especially useful
 CC for preventing redeposition of colorant during stonewashing, and for
 CC processing of textiles where cellulose breakdown is required. The new

CC truncated enzymes show reduced redeposition of dye compared to using
 CC non-truncated cellulase compositions.

SQ Sequence 1751 AA;

Query Match 21.3%; Score 1524; DB 20; Length 1751;
 Best Local Similarity 53.5%; Pred. No. 2.1e-83;
 Matches 289; Conservative 76; Mismatches 151; Indels 24; Gaps 6;

QY 32 PTPVTVHGQT-----AKTVTKVDTSKDRKPIPIYIYGTNQELAGDENLTARRLGGNRM 86
 Db 1215 ptpvtvtpptpavtpdvkisdtsgrtkispyiyganqdiqgvvh-parrlgnrnt 1273
 QY 87 GYNWENNNSAGSDWQSSDLYCDNAGLTAKACEKPGAVATSFHQSLKQGTYSLVTL 146
 Db 1274 gynwnnmsnagsdwyhssddymcylmgltgndknvpavavvskfheqslkqnsyatlq 1333
 QY 147 MAGYVAKDNGSVQSEKAPSAARNVNAKNAPFOLQDPLKDNVYVADSEVNFVKYK 206
 Db 1334 mgyvvakdngsvsesetapsrwaevkfkkdgalslqpdvndnyymdefinylknkg 1393
 QY 207 VASTKTGVKCYSLDNEPALMSHTHPRIHGEKVGAKELVDRSVLSKAAKAVDAGAEIFGP 266
 Db 1394 rsssatgikgyilidnepdlwfttthpripqkvctscselinksvlakviktldpdaeifgp 1453
 QY 267 VLYGFGAYKDLQAPDWNVSKGNYSWFVDYLLDOMRLSSQAEGRLLDVFVHWYPEAMG 326
 Db 1454 asygfvgylclqapdwngvqgnhrwflswyleqmkksadsfgkrlldvldihwypeaqv 1513
 QY 327 GGIRITNEVGNDETCK-----ARMQAPRTLWDPTYK-----EDSWIAQWNSSEFLPL 372
 Db 1514 ggvricfd-genstsrdaivarmqaprtlwdptyktqkgitagenswinqwfpaylpl 1572
 QY 373 LPRLKQSVDKYYPGCTKLALTEYSGGENDISGGIAMDVLGILKNDNYVMANYWLKDDGA 432
 Db 1573 lpnikadidkypgctklaitefdyggkdhisggialadvlfgfygvymaarw---gds 1629
 QY 433 NNYVSAAYKLYRNYDGKSTFGDISVHAOTSDIVNSSVHASVTDASYKELHLVNMKNSMD 492
 Db 1630 gsyaaqayniynydgkgyrystcvsaetdvempvyasiegeddstcvhlllnrnyd 1689
 QY 493 SAFDAQFDLSGETTYGSGKVMGFDKNSOIKEAAPITQISGNRFTYVPLTAHYHVLTA 552
 Db 1690 rklkaeikmntrvytggelygfdstssqirkmgvlsnqntitievplntvyhivlts 1749

RESULT 7

AAY54135
 ID AAY54135 standard; Protein: 496 AA.

AC AAY54135;

DT 27-MAR-2000 (first entry)

DE Amino acid sequence of a *Bacillus* sp. mannanase enzyme.

KW Mannanase; mannan endo-1,4-beta-mannosidase; beta-mannanase;
 KW endo-1,4-mannanase; galactomannan; 1,4-beta-D-mannosidic linkage;
 KW mannan; galactomannan; glucomannan; galactoglucomannan; cellulosic fibre;
 KW synthetic fibre; yarn; fabric; printing paste; thickener; viscosity;
 KW plant material degradation; recycled waste paper; paper making pulp;
 KW guar; locust bean gum; mannan-containing food; coffee extract;
 KW cleaning composition; machine washing; hard-surface cleaner;
 KW dishwashing; oral; dental; contact lens; body-care composition;
 KW fabric softener; oil well drilling; subterranean formation fracture.

OS *Bacillus* sp.

FH Key Location/Qualifiers

FT Peptide 1..37

FT /note= "signal peptide"

FT Domain 38..165

ID	AA113492 Standard; Protein; 1420 AA.
XX	

Qy	1047	VSPLELVRYMV-----WGFLERRQTK-----QSLEHLKKV-----YNH	1080
Db	442	mistlgedgypvllpqnkgdywgsnliankmillgysdlsqndyikannrmgmdyfmng	501
Qy	1081	PNVITLEKPTNLKTYGITEQPSVPGSPFTLNAAGETAKVLSWTFASANAASAEVVKSTVE	1140
Db	502	nnamrlsyi-tgygqyyetdthd-----rwagkyqsgvpy--pkgwls	542
Qy	1141	NGAFATVSSDVWGSSYTDYNTADTYYVQVAKN-----DACQTVSNT-----	1184
Db	543	ggpntvind---satrtgmpaaksy---apknatpadawvskentlnwnaplwlwisky	594
Qy	1185	-----ASAAPKAD-----TQQTPTGLVLQYRTADTIVNNDNHLNHPHF	1220
Db	595	iqdhreplgganqnppdpqppepepepnpppsgdlaavkyktgdtnaadnqkphf	654
Qy	1221	QILNKGTISVPINEUKIRYYTIDGRQTFNCDYAVLUSCKLNGKLVKMDKAATGADYY	1280
Db	655	nivnkgaaavpiseisrlryftadgndqlqncddwamvgcsnlnngatvknmpkanady	714
Qy	1281	LEVFSNMSDAGVLASGGSTGGIOTRTHKADWSNYESDDYSYKGTQTSFDDHTRATLYHNG	1340
Db	715	leitfkaagdsiqpagkrgisrtrhagnwanlnhesndysfdpkkttayanwervtlyhng	774
Qy	1341	VLVWGTE 1347	
Db	775	tlvfigne 781	

RESULT 13
AAV13492
ID AAV13492 standard; Protein; 1426 AA.
XX

CC non-truncated cellulase compositions.

AA	Sequence	616 AA;
SQ		

Query Match	5.28;	Score 375;	DB 20;	Length 616;
-------------	-------	------------	--------	-------------

Best Local Similarity 44.7%; Pred. No. 1.9e-14;

Matches 67; Conservative 35; Mismatches 48; Indels 0; Gaps 0;

QY 1198 TGLVLQYRTADTNVNDNHLNPHFQILNKGTSVPINELKIRYYTIDGDRQTFNCDYAV 1257

Db 3 sqkvlyknnetsastqsrpwfkiivnggssvdlrsvkirvwyvtvdgdkpqsavcdwaq 62

QV 1258 LSCSKLNGKLVKMDKAATGADYYLEVSNFSDAGVLASGGSTGGIQTIRIHKADWSNYESD;1317

63 i qasnytfnfyk|ssgvsqadv|evafssgag|apqkdtad|avrfnkndwsnynqad 122

QY 1318 DYSYKGTQTSFDDHTKATLYHNGVLVWGTE 1347

```

123 dws\lasmt\nvg\enakv\tlvyd\g\lvwqge 152
db
|:|: : |: : | |: : ||| : |||| |

```

RESULT 15

AAR42122

ID	AAR42122 standard; Protein; 499 AA.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100
101	101
102	102
103	103
104	104
105	105
106	106
107	107
108	108
109	109
110	110
111	111
112	112
113	113
114	114
115	115
116	116
117	117
118	118
119	119
120	120
121	121
122	122
123	123
124	124
125	125
126	126
127	127
128	128
129	129
130	130
131	131
132	132
133	133
134	134
135	135
136	136
137	137
138	138
139	139
140	140
141	141
142	142
143	143
144	144
145	145
146	146
147	147
148	148
149	149
150	150
151	151
152	152
153	153
154	154
155	155
156	156
157	157
158	158
159	159
160	160
161	161
162	162
163	163
164	164
165	165
166	166
167	167
168	168
169	169
170	170
171	

XX
AC

XX
AC
AAK42122;

DT 27-APR-1994 (f

XX

DE NK-1 cellula

XX

KW Cellulase; p
yy

XX QS Bacillus NA

XX
33
EN STATION

PN JP05236969-A.

XX

PD 17-SEP-1993.

XX	28-FEB-1992;	92JP-0075883.	
PF			
XX			
XX	28-FEB-1992;	92JP-0075883.	
PR			
XX			
PA	(BEPP/) BEPPU T.		
XX	WPI; 1993-330585/42.		
DR	N-PSDB; AAQ49820.		
XX			
PT	Changing the pH-dependence of cellulase enzymatic activity - by		
PT	changing base sequence of cellulose-producing gene of <i>Bacillus</i>		
PT	microbe to base sequence coding asparagine and serine residues at		
PT	specified aminoacid sites		

PS Disclosure: Fig 1: 9pp: Japanese:

XX Sequences (AAQ49818-19) consist of two synthetic fragments which are
CC used to induce a mutation within the cellulase NK-1 gene. The gene
CC is shown in sequence (AAQ49820). The modified NK-1 gene shows a
CC change in pH dependence.

Sequence 499 AA:

Query Match

Query match 4.56; score 552.5; DB 1
Best Local Similarity 35.18; Pred. No. 3.3e-13;

Best local similarity	55.19; Acc: 80.33	55.19; Acc: 80.33	55.19; Acc: 80.33
Matches	80; Conservative	36; Mismatches	95; Indels
			17; Gaps
			7;

QV 1121 LSWTASANAASYEVKRSTVENGAFAFATVASDVYGSSYTDTAVTADTTYQ - VVAKNDAGQ 1179

pb 286 isw-vnwn|sd|kaessal|kogaskt-----ggw|td|tasat|fvenil|ankdstk 337

QV 1180 TVSNTASAAPKADTOOPTTGLVLOYRTADTNVNDNHLNPHFOILNKGTTISVPINELKIRY 1239

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:49:37 ; Search time 35.21 seconds
(without alignments)
936.511 Million cell updates/sec

Title: US-09-784-554b-4

Perfect score: 7165

Sequence: 1 MKAKNSSISWRSKWLPPV.....HTKATLYHNGVLWGVTETAN 1350

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1816.5	25.4	700	2	US-07-862-588B-2
2	1797	25.1	617	1	US-08-361-920-29
3	1797	25.1	617	1	US-08-479-939-29
4	1797	25.1	617	1	US-08-483-432-29
5	1524	21.3	1751	4	US-09-136-574A-44
6	550.5	7.7	551	2	US-09-033-537A-1
7	423.5	5.9	1426	4	US-09-136-574A-43
8	375	5.2	616	4	US-09-136-574A-47
9	336.5	4.7	493	4	US-09-198-956-10
10	336.5	4.7	493	4	US-09-198-955A-12
11	320	4.5	531	2	US-07-862-588B-7
12	303	4.2	167	5	PCT-US95-13813-9
13	276	3.9	336	3	US-09-095-163-2
14	229	3.2	162	1	US-08-048-164A-2
15	229	3.2	162	1	US-08-460-462-2
16	229	3.2	162	1	US-08-460-457-2
17	229	3.2	162	1	US-08-460-458-2
18	229	3.2	162	2	US-08-460-455-2
19	229	3.2	162	2	US-08-330-394A-2
20	229	3.2	163	3	US-09-006-636-7
21	229	3.2	163	4	US-09-006-632-7
22	223	3.1	154	2	US-08-330-394A-29
23	223	3.1	156	2	US-08-330-394A-22
24	223	3.1	382	4	US-09-277-716-22
25	200	2.8	2411	4	US-09-268-347-36
26	197	2.7	2353	4	US-09-377-155-33
27	197	2.7	2353	4	US-08-913-942-4

28	197	2.7	2353	4	US-09-669-974-33	Sequence 33, Appl
29	189.5	2.6	2354	4	US-09-268-347-47	Sequence 47, Appl
30	182.5	2.5	1912	1	US-08-409-995-4	Sequence 4, Appl
31	182.5	2.5	1912	3	US-08-685-467-4	Sequence 4, Appl
32	174.5	2.4	1222	2	US-08-682-517-15	Sequence 15, Appl
33	174.5	2.4	1252	2	US-08-682-517-9	Sequence 9, Appl
34	170.5	2.4	1002	4	US-09-268-347-24	Sequence 24, Appl
35	161.5	2.3	592	1	US-08-217-327-8	Sequence 8, Appl
36	160	2.2	866	1	US-08-386-727-8	Sequence 8, Appl
37	160	2.2	866	2	US-08-600-452A-8	Sequence 8, Appl
38	157.5	2.2	1810	5	PCT-US95-11684-4	Sequence 4, Appl
39	156.5	2.2	872	2	US-08-387-942C-5	Sequence 5, Appl
40	156	2.2	1375	4	US-09-210-361-4	Sequence 4, Appl
41	156	2.2	1732	2	US-08-570-311-10	Sequence 10, Appl
42	156	2.2	1732	2	US-08-353-485-10	Sequence 10, Appl
43	154	2.1	1169	1	US-08-542-921-2	Sequence 2, Appl
44	154	2.1	1169	2	US-08-880-685-2	Sequence 2, Appl
45	154	2.1	1169	2	US-08-880-684-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnae
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-862-588B-2

Query Match 25.4%; Score 1816.5; DB 2; Length 700;
Best Local Similarity 52.3%; Pred. No. 1.2e-128;

Qv 165 APSARWNEVVNAKNAPFOLOPDLKDNOVYADEFVNELVKKYGVASTKTGVKGYSLDNEPA 224

; FILING DATE: 08-MAY-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lambiris, Elias J.
 ; REGISTRATION NUMBER: 33,728
 ; REFERENCE/DOCKET NUMBER: 3435.204-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-867-0298
 ; INFORMATION FOR SEQ ID NO: 29:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 617 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-483-432-29

Query Match 25.1%; Score 1797; DB 1; Length 617;
 Best Local Similarity 55.2%; Pred. No. 2.9e-127;
 Matches 349; Conservative 73; Mismatches 146; Indels 64; Gaps 6;
 QY 45 VTIKVDTSKDKRPISPIYGTGNQELAGDENLTARLGGNRMGTGYNNWNNMAGSDWQWS 104
 Db 28 VFTTINTQSERAASPNYGTQDLSGTENNSSRRLLGNRLTGYNNWNNASSAGRWLHY 87
 QY 105 SDSYLDNAGLTAKCEKPGAVATSEHDOSLKQGTYSLVTLPMAGYVAKDNGSGVQSEK 164
 Db 88 SDDFLCGNGVDPDCCDKPGAVVTAHDKSLGAYSLVTLQAMAGYVSRDKNGPVDSE 147
 QY 165 APSARWNEVNAKNAPFQLOLDKNQVYADEFNFLVKKYGVASTKTGYKGYSLDNEPA 224
 Db 148 APSRWDKVEAKNAPESLQPLDNGQVYNDVEFNFLVKNYGNASTGTGKAYSLDNEPA 207
 QY 225 LWSHTPHRIHEKVGAKELVDKRSVLSKAAKAVDAGAEIFGPVLYGFGAYKDLQATPDWN 284
 Db 208 LWSETHPHRIHQALAEVAKSADLSKAVKNDVPHAEIFGPALYGFAYLSLQADPDW 267
 QY 285 SVKGNYSWFDYLDQMLRSQAEGKRLDVFVHWYPEAMGGGIRIT-NEVNDKTKA 343
 Db 268 SLQNGYSWFDYLDQMKNAHTQNGKRLDVLVHWYPEAQQGGQRIYFGGAGNIDTQKA 327
 QY 344 RQAPRLMDPTTKEDSWIAQWNESEFLLPLRLKQSDYKDYPGTKALTEYSYGGENDIS 403
 Db 328 RVQAPRLMDPAYQEDSWIGTWESSYLPLPKLQSSITQYTPGKALTEFSYGGNDIS 387
 QY 404 GGIAADVGLTLCKNDYMYANKYKLDGANNYSAAKLYRNDYDGKSSFTGDISVHAQTS 463
 Db 388 GGIAADVGLTLCKNDYMYANKYKLDGANNYSAAKLYRNDYDGKSSFTGDISVHAQTS 446
 QY 464 DIVNSSVHASVTDAKYKELHLVNNKSMDSAFDAQDLGSETTYGSGYKVGFDKNSQIK 523
 Db 447 DTENSSVYASVTDEENSELHLVNNKFNDDPINATFQSGDKTYTSGRWGFPDQSGDIT 506
 QY 524 EAPITQISGNRTYTPPLTAYHIVLTAGNDTPVENPESFALRAEAGDKSYLDASSG 583
 Db 507 EQAATINNNQETTYLPLLSAYHIVLKADSTEPVISE----- 544
 QY 584 VVGYSVQRATYENGPPAAVASNLVETSYDTNTWNTGTSYTKITAKTKGTGSASVNLKAV 643
 Db 545 -----IPSSSTSSPVNQTS-----TSTTSTSTSSPPVQ--- 574
 QY 644 PRAPVDPDRYAEADGTLKGTVVESGTFSG 675
 Db 575 PTPP-----SCTAERWAQCGNGWSG 596

RESULT 5

; US-09-136-574A-44
 ; Sequence 44, Application US/09136574A
 ; Patent No. 6294366
 ; GENERAL INFORMATION:
 ; APPLICANT: Farrington, Graham K.
 ; Anderson, Paige

; Gibbs, Moreland
 ; Bergquist, Peter
 ; Daniels, Roy
 ; Morgan, Hugh W.
 ; Williams, Diane P.
 ; TITLE OF INVENTION: Compositions and Methods for
 ; Treating Cellulose Containing Fabrics Using Truncated
 ; Cellulase Enzyme Compositions
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Howson and Howson
 ; STREET: Spring House Corporate Center, P.O. Box 457
 ; CITY: Spring House
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19477
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/136,574A
 ; FILING DATE: 19-Aug-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/932,571
 ; FILING DATE: September 19, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bak, Mary E.
 ; REGISTRATION NUMBER: 31,215
 ; REFERENCE/DOCKET NUMBER: 1997US001/CIP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-540-9200
 ; TELEFAX: 215-540-5818
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 44:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1751 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 ; US-09-136-574A-44

Query Match 21.3%; Score 1524; DB 4; Length 1751;
 Best Local Similarity 53.5%; Pred. No. 6.9e-106;
 Matches 289; Conservative 76; Mismatches 151; Indels 24; Gaps 6;
 QY 32 PPTTVVHGQT-----AKTVTIKVDTSKDRKPTSPYIYGTNQLAGDENLTARRLGGNRM 86
 Db 1215 PPTTVVTPPTPAVTPDVKISIDTSRGRTKISPIYGANQIOGVVH-PARRLGGNRLT 1273
 QY 87 GYNWNNMNSAGSDWQSDSYLDNAGLTAKCEKPGAVATSEHDOSLKQGTYSLVTLPL 146
 Db 1274 GYNWNNMNSAGSDWYHSSDDYMCYIMGITGNDKNVPAAVVSKFEHQSIKQNASITLQ 1333
 QY 147 MAGYVAKDNGSGVQSEKAPSAARNVNNVNAKNAPFQLOLDKNQVYADEFNFLVKKY 206
 Db 1334 MVGYVAKDNGTGVSESETAPSRWAEVKKKDGALSGALSDPDVNDYVYNDDEFINLYNKY 1393
 QY 207 VASTKTGYKGYSLDNEPALWSHTPHRIHEKVGAKELVDKRSVLSKAAKAVDAGAEIFGP 266
 Db 1394 RSSSATGKGYLDNEPDLPWETTHPHRIHQVKTCSSELINKSVELAKVIKTLDPDAEIFGP 1453
 QY 267 VLYGFGAYKDLQATPDWNVSKGYSWFDYLDQMLRSQAEGKRLDVFVHWYPEAMG 326
 Db 1454 ASYGFVGYLTLDQADPDWNQVKNHRWFLSWYLEQMKASDSFGKRLDVLDIHWYPEAQV 1513
 QY 327 GGIRITNEVGNDKTK-----ARMOAPRLMDPTTK-----EDSWTAQWNESEFLPL 372
 Db 1514 GGVRLCFD-GENSTSRDVAIARMOAPRLMDPTTKTKQGOITAGENSWINQWPFYPLPL 1572

;; FILING DATE: September 19, 1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: 1997US001/CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-540-9200
;; TELEFAX: 215-540-5818
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 43:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1426 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: No. 6294366e
;; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43

Query Match 5.9%; Score 423.5; DB 4; Length 1426;
Best Local Similarity 27.0%; Pred. No. 6.2e-23;
Matches 127; Conservative 67; Mismatches 171; Indels 105; Gaps 12;

QY 967 DRLTNVHKLNL-----IWNWSVAPDWPYCGDEYVDILSPDS-----YPOAGDYSP 1012
DB 108 DTFVNFANTNIGIRGTLVWHNTDPMFFRDSSGQMLSKDALLARLKQYIDVVGRYKG 167
QY 1013 QIAKYEDLVTLGDKLKVCHERKRTY---PQPGSDE-----GVSSPL 1051
DB 168 KVIW-DVNEAIDESPQDGYRRSTWQICGPEVIEKAFIWAHEADPAKLFNDYNTET 226
QY 1052 ELVRYVWVGFLERRQTKQSLHLKKVYNHPV-----ITLEKL 1089
DB 227 STKRDFYINMYKLNKSGVPIHGIMQSHINVMWPSVSEIENSIKLFSSIPGIEIHTEL 286
QY 1090 PTNLKTYGITEQSPVPGSFTLNAAGT-----AKVKLSWTASA 1127
DB 287 DMSLYNNGSNENYTPQDQLLORQAQYKIDFTMLRKYKIVTCVTFWGLKDDYSLNS 346
QY 1128 NAA-----SYEVKSTVENGAFAFVASDVYGSSTYDTAVTADTYYYQVVAKNDAQ 1179
DB 347 SKRDWPLLFDDYSK-----PAYWSVIEAGASASPSPTVATPT-----PTPTP 392
QY 1180 TVSNSTAAAKADQOQPTT--GLVLOYRTADTNVNDHNLPHQILNKGTISVPINELKI 1237
DB 393 TVTVATPTP--PTGTPGSGGLKLVLYKNNETSASTGSRPWFKIVNGSSSDVLSRVKI 451
QY 1238 RYYTIDGDEQTFNCDDYAVLSCGLKGLVKMDKAATGADYILEVSEFNSDAGVLASGGS 1297
DB 452 RYWTVDGDKPQSAVCDDMAQIGASNTFTFVKLSSGSGADYILEVGFSSGAGQLQPGKD 511
QY 1298 TGGTQTRIHKADSNYNESDDYKGTQTSFDDHTKATLYHNGVLVWGTE 1347
DB 512 AGDIQVRFNKNDWSYNOADWSLQSMITDYGNAKVTLYVDGVLVWGQE 561

RESULT 8
US-09-136-574A-47
; Sequence 47, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Howson and Howson
;; STREET: Spring House Corporate Center, P.O. Box 457
;; CITY: Spring House
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19477
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/136,574A
;; FILING DATE: 19-Aug-1998
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/932,571
;; FILING DATE: September 19, 1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: 1997US001/CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-540-9200
;; TELEFAX: 215-540-5818
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 47:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 616 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

Query Match 5.2%; Score 375; DB 4; Length 616;
Best Local Similarity 44.7%; Pred. No. 7e-20;
Matches 67; Conservative 35; Mismatches 48; Indels 0; Gaps 0;

QY 1198 TGLVLOYRTADTNVNDHNLPHQILNKGTISVPINELKIRYYTIDGDRQTFNCDDYAV 1257
DB 3 SGVKVLYKNNETSASTGSRPWFKIVNGSSSDVLSRVKIYWTVDGDKPQSAVCDDMAQ 62
QY 1258 LSCSKLNGKLVKMDKAATGADYILEVSEFNSDAGVLASGSGTGTQTRIHKADSNYNESD 1317
DB 63 IGASNTFTFVKLSSGSGADYILEVGFSSGAGQLQPGKDITGDIQVRFNKNDWSYNOAD 122
QY 1318 DYSYKGTQTSFDDHTKATLYHNGVLVWGTE 1347
DB 123 DMSWLQSMITDYGNAKVTLYVDGVLVWGQE 152

RESULT 9
US-09-198-956-10
; Sequence 10, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377,200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02

REFERENCE/DOCKET NUMBER: 5609.000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-163-2

Query Match 3.9%; Score 276; DB 3; Length 336;
Best Local Similarity 27.3%; Pred. No. 7.6e-13;
Matches 87; Conservative 52; Mismatches 118; Indels 62; Gaps 10;

QY 781 TKTLINPNATVEAKALMNYLV---DQYGRNMLSG-----QEDMPEIDWLQANVGGKP 829
D 2 TVSPVNPNAQQTTKTVNNWLAHLNPTNRNVLGSAFGGYSHTFSAEADRIKRSATGQSP 61
QY 830 AIALDLIDYSPSRAEHLSSSTETAKAIEWDKOGGIVTFAWHNNAKGLIDTQGKEWWRG 889
D 62 AI-----YGCYARCWLETAIEDSIDVSCNGDLMSYKNGGIPQISHLANPAFQSG 114
QY 890 FYADSTTFDIYAMNHPESEDYKLLINDIVIAQQLKKLQDAKVPVFLRPLHLEAEKGWFW 949
D 115 HFKPTITDQYKNLDSATAEGRKRLNAMLKSIADGLQELNENQGVVFLRPLHEMNGEWF 174
QY 950 WG-----AKPEPV---KKLYILMHDRLTNVHKLNLWVNSVA-----PDWYPGDEV 996
D 175 WGLFSYNKONERISLYKQYKIIYHYMTDTRGLDHLIWIYSPDNRDKFTDYPGASIV 234
QY 997 DILSFDSPQAGDYSPOIAKYEDIVTLGK-----DKKLVCHEKRTYTPG 1040
D 235 DIVGLDAYFQDA-Y\$--INGYDQLTALNKPFAFTEVGPQTANGSFDYSLFNAIKQYP- 290
QY 1041 PGSDEGVSSPLELVRYMW 1059
D 291 -----KTIYFLAW 298

RESULT 14
US-08-048-164A-2
Sequence 2, Application US/08048164A
Patent No. 5496934
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Doi, Roy H.
TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/048,164A
FILING DATE: 14-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mirostock, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7809-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-048-164A-2

Query Match 3.2%; Score 229; DB 1; Length 162;
Best Local Similarity 34.0%; Pred. No. 8e-10;
Matches 54; Conservative 30; Mismatches 65; Indels 10; Gaps 3;

QY 1197 TTGLVLOYRTADTNVNDNLNPHFOILNKGITISVPINELKIRYYYYTIDGREQTFNCDYA 1256
D 3 TSSMSVEFYNSKNKSAQTNSITPIIKITNTSDNLNDVKRYYYTSDGTGQIFWCDHA 62
QY 1257 -----VLSCSKINGKLVMKMDKAATGA-DYYLEVSFNSDAGVLASGGSTGGTQTRHK 1307
D 63 GALLGNSYVDNTSKVTANFVKETASPTSTYDTYVFEFGFASGAATLKGQFI-TIOGRITK 121
QY 1308 ADWSNYSNDDYSYKGTQTSFDDHTKATLYHNGVLVWGT 1346
D 122 SDMSNYTQTDNSYDASSSTPVVNPVKVTGYIGGAKVLGT 160

RESULT 15
US-08-460-462-2
Sequence 2, Application US/08460462
Patent No. 5670623
GENERAL INFORMATION:
APPLICANT: Shoseyov, Oded
APPLICANT: Shpiegl, Itai
APPLICANT: Goldstein, Marc A.
APPLICANT: Doi, Roy H.
TITLE OF INVENTION: METHODS OF USE OF CELLULOSE BINDING DOMAIN PROTEINS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,462
FILING DATE: concurrently herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,164
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mirostock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7809-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:48:51 ; Search time 59 Seconds
(without alignments)
2198.654 Million cell updates/sec

Title: US-09-784-554B-4

Perfect score: 7165

Sequence: 1 MKANKSSINSKRSKWLPPV.....HTKATLYHNGVLVWGHTETAN 1350

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR71.*
1: Pirl.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1797.5	25.1	700	2 B41897	cellulase (EC 3.2.
2	1534	21.4	606	2 H97012	hypothetical prote
3	1454.5	20.3	1331	2 A48954	mannan endo-1,4-be
4	822.5	11.5	1021	2 T10748	mannan endo-1,4-be
5	801.5	11.2	591	2 JC7657	mannan endo-1,4-be
6	439	6.1	1779	2 T31085	xylanase - Caldice
7	397.5	5.5	915	2 A43802	cellulase (EC 3.2.
8	397	5.5	1039	2 S02711	cellulase (EC 3.2.
9	391.5	5.5	1711	2 T31337	1,4-beta-glucanase
10	372	5.2	1742	2 T17120	cellulase (EC 3.2.
11	369.5	5.2	535	2 A41375	cellulase (EC 3.2.
12	356.5	5.0	508	2 A26874	cellulase (EC 3.2.
13	352.5	4.9	499	2 A27198	cellulase (EC 3.2.
14	351	4.9	145	2 A41897	cellulase homolog
15	348.5	4.9	499	2 JN0111	cellulase (EC 3.2.
16	347.5	4.8	508	2 G69593	cellulase (EC 3.2.
17	339.5	4.7	1854	2 S36859	cipa protein - Clo
18	328	4.6	986	2 S12021	thermoactive cellu
19	319	4.5	879	2 A47704	endoglucanase I (E
20	291	4.1	1483	2 C97012	probably celluloso
21	289.5	4.0	486	2 I40548	bifunctional cellu
22	283	3.9	362	2 H69785	mannan endo-1,4-be
23	279.5	3.9	505	2 S39962	endoglucanase - Er
24	277	3.9	504	2 S54744	cellulase (EC 3.2.
25	275.5	3.8	1475	2 A44765	alpha-amylase (EC
26	270.5	3.8	586	2 PC6006	scaffolding protei
27	268	3.7	513	2 A37219	mannan endo-1,4-be
28	260.5	3.6	1230	2 S47466	cellulose 1,4-beta
29	254.5	3.6	1162	2 T30433	scaffolding protei

30 247 3-4 360 2 S60268 beta-mannosidase (
31 247 3-4 419 2 S53374 mannanase A - pseu
32 231.5 3.2 1203 2 S27545 pullulanase - Ther
33 226 3.2 1848 2 A44140 cellulose-binding
34 216 3.0 1481 2 S28669 amylopululanase p
35 207.5 2.9 1282 2 JC4393 microbial collagen
36 207.5 2.9 1684 2 S10789 anylase A-180 - al
37 207.5 2.9 2044 2 AB1180 spoIID-like domain
38 206 2.9 1042 2 A97209 cell wall-associat
39 206 2.9 2167 2 AF1489 probable peptidogl
40 205.5 2.9 1993 2 AF1450 beta-mannanase Man
41 204.5 2.9 505 2 C96966 probable S-layer p
42 203.5 2.8 1939 2 D97316 glycosidase homolo
43 200.5 2.8 1090 2 AG1749 titin, cardiac mus
44 199.5 2.8 26926 1 I38344 titin - rabbit (fr
45 198.5 2.8 6605 2 S20901

ALIGNMENTS

RESULT 1

B41897

cellulase (EC 3.2.1.4) - Bacillus lautus

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Bacillus lautus

C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 15-Oct-1999

C:Accession: B41897; S27499

R:Hansen, C.K.; Diderichsen, B.; Jorgensen, P.L.

J. Bacteriol. 174, 3522-3531, 1992

A:Title: cella from Bacillus lautus PL236 encodes a novel cellulose-binding endo-beta-

A:Reference number: A41897; MUID:92276330

A:Accession: B41897

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-700 <HAN>

A:Cross-references: EMBL:M76588; NID:g142661; PIDN:AAA22303.1; PID:g142663

A:Experimental source: PL236

A:Note: sequence extracted from NCBI backbone (NCBIP:104605)

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as

A:Pathway: cellulose degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match

Best Local Similarity 25.1%; Score 1797.5; DB 2; Length 700;

Matches 368; Conservative 86; Mismatches 183; Indels 71; Gaps 13;

QY 11 SKRSKWLPPVVMACITIVGGALPPTVVGHTAKTIVTKVDTSKDRKPISPIYGTNOELA 70

Db 3 TRORRKLRFVSAALAVSLTWTVPMPASVNA-AASDVTFTINTQSERAAISPNIYGTNODLS 61

QY 71 GDENLTRARRLGGNRMGTGYNNENMNSAGSDMQSSSYLCDNAGLTAKCEKPGAVATSF 130

Db 62 GTENSSRRLLGGNRLTGYNNENMNSAGSDMQSSSYLCDNAGLTAKCEKPGAVATSF 121

QY 131 HDOSLQKQYSLVTLPMAGYVAKDNGSVQSEKAPESARNEVNVNNAKNAFLOLPDLKDN 190

Db 122 HDRSLNGAYSIYTLQMGYVSRDKNGPVDSETPSPRWDKVEFPAKNAPFSLQPHLNDG 181

QY 191 QYVADFVNLVKYGVASTKTGKGYSLDNEPALMSHSTPHRIHGEKVGAKELVDRSVSL 250

Db 182 QVTMDEVNVLVRYGNASTSTGIKAYSLDNEPALMSETHPRTHPEQLQAAELVAKSIDL 241

QY 251 SKAAKAVDAGAEITGPPVLYGFGAYKDLQTPADWNVSKGVNYSWFVYLLDQMRSSQAEK 310

Db 242 SKAVKNVDPAEITGPPVLYGFGAYKDLQTPADWNVSKGVNYSWFVYLLDQMRSSQAEK 301

QY 311 RLDDVDFVHVHPAMGGGIRIT-NEVGNDETKKARQAPETLMDPTTKEDSWIAQNSEF 369

Db 302 RLDDVDFVHVHPAMGGGIRIT-NEVGNDETKKARQAPETLMDPTTKEDSWIAQNSEF 361

QY 370 LPPLRLKOSVDKYPPGTKLALTEYSYGGENDISGGTAMADVILGLCKNDVWANYWKLK 429
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 362 LPIELPKLOSQTQTYPGTKLAITESSSYGDNHISGGIATADALGPFGYCVYAANYWQTE 421
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 430 DGCANNVSAAVKLYRNYDGRKSSTFDDLSIVHAQTSIDVNSSVIASVTDSASYKEHLVLWNKK 489
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 422 DN-TDYTSAALKLRNLDGNKGSGSLIKYDAATSDFSENSVASVTDEENSELHLIVLNK 480
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 490 SMDSADFDAQFDLSGETTYGSKGVMGFDKNSSQIKEAAPITQISGNRFVTVPPLTAYHV 549
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 481 NFDDPINATEFOLS GDKTYTGSRVWGFEQDTGS DITEQAATINNNOQFYTTLPPLSAYHV 540
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 550 LTAGNDTPVENPESFAIRAEAGDGKSIYLSDASSGVVGYSVRATYENGEPFAAVASNLVET 609
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 541 LKAUSTEPTV-----NSDLV-----VQKGDGRNNATDNIKP 572
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 610 SYTDINVTN-GTS-----YYIKTAKTGTGSASNVLKAVPRAPDVDPD---RYE 655
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 573 HF--NQNKNGSPVDSLSSLTRYEF-----FKDSSAAN--GWIDNAKLGSGNIQISFG 622
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 656 AEDCTLAGTVVE---SSGTCFSGTGYV-----TNFHAGD 687
| |: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 623 NHNGADSOTVAELGFFSGAGSIAEGQGSEIQLRMSKADWSNFNEAND 670

RESULT 2

H97012
hypothetical protein CAC0915 [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97012
F:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,
J.; Daly, M.J.; Bennett, G.N.; Koehn, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacillus
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: H97012
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-606 <XUR>
A:Cross-references: GB:AE001437; PIDN:AAK78891.1; PID:g15023815; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0915

```

Query Match      21.4%: Score 1534; DB 2; Length 606;
Best Local Similarity 51.3%: Pred. No 7.3e-80;
Matches 291; Conservative 85; Mismatches 175; Indels 16; Gaps 7

```

```

QY 367 SEFLPLLPKLSVDKYYPGPTKLALTEYSYGENDISIGIAMADVLGITLGNKVYMANVW 426
    |||:|:| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 358 KWGLPLIPKVKSSIDIKYFPGTKLFSSEYNCGEDHISGGIAQADALGVFGKYGYFYATYW 417
    |||:|:| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 427 KLKDGANNVVSAAYLKRYNCKSGTSGDISVHAQTSIVNVSSVHASVTDSASYKEHLVW 486
    |||:|:| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 418 ECNSDKNNYVQSFAFNLYNNYDGNNSKYGDTDVCKVCTSDINNSSVTASVTSNDGNKMMDIIV 477
    |||:|:| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 487 MNKSMDSAFDAQDFLDGETTYGSKRVGWFDKNSSOIKEAAPITQISGRNFRTYVPPLTAY 546
    |||:|:| :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 478 MNKNYTDSINFNVSNKKNTSGOVWGFDSSNNITRDDVSSISGNKFYKIPALTAV 537
    |||:|:| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 547 HVLVTAGNDTPVENPESFALRAEAGDG 573
    |||:|:| :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 538 HVLTTAQKSSVKGDVN-----GDG 557
    |||:|:| :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

RESULT * 3
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N:Alternate names: beta-mannanase
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A48954; B43745
R:Gibbs, M.D.; Saul, D.O.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol.: 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidoma
A:Reference number: A48954; MUID:93119139
A:Accession: A48954
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: GB:L01257; NID:g144290; PIDN:AAA71887.1; PID:g144291
R:Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol.: 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene cod
A:Reference number: A43745; MUID:91247819
A:Accession: B43745
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQRQ' <LUE>
A:Cross-references: EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PID:g144294
A>Note: the authors translated the codon CAC for residue 262 as Glu
A>Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

```

[illegible]

RESULT 6
T31085
xylanase - Caldicellulosiruptor sp.
C:Species: Caldicellulosiruptor sp.
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31085
R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.
submitted to the EMBL Data Library, December 1997
A:Description: Family 10 and 11 xylanase genes from Caldicellulosiruptor sp. Rt69B.1.
A:Reference number: 220972
A:Accession: T31085
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1779 <MOR>
A:Cross-references: EMBL:AF036924; NID:g2760905; PID:g2760909; PIDN:AAB95326.1
C:Genetics:
A:Note: xync

Query Match 6.1%; Score 439; DB 2; Length 1779;
Best Local Similarity 20.4%; Pred. No. 1.6e-16;
Matches 304; Conservative 165; Mismatches 441; Indels 580; Gaps 68;

QY 10 WSKRSKWLPLVVMACIIIVGGALPTPTVVGOTAK--TVTIKVDTSKDR---KPISPYIY- 63
DB 200 WQARSGVTVPDPSVAYNGKY--SLYVSGRTSNHGAQIPVDITLGGKVKYKISVWVYQ 257
QY 64 --GTNOELAGDENLT--ARRLGGNRTGY--WNENNSNAGSDWQSSDSY----- 108
DB 258 NSGSTQKM---SLTMORREATDPSTSYENLIYNRDVPSTNWTVEPSGYSIPAGVTYSEL 313
QY 109 -----LCDNAGLTKACEKPG-----AVATSFHDQSLKQG 138
DB 314 LLYVEAQANAFVDDDKITDLSKLAEPWEIPSLIEKYKDYFKVGVALSY--KSTASD 371
QY 139 TYSILVTPMAGYVADKNGSVQSEKAPSARWNEVNAKNAPOLOQLDKDNQVYADEFV 198
DB 372 TEKKMVL-----KHENSITAGNEMKPS-----NYNFSKADFEV 411
QY 199 NFLVKYGVASTKGVKGYSLDNEPALWSHTPHRIGEKVGAKELVDRSVLSKAAKAVD 258
DB 412 NF-----ATSNNTAIRGHTL-----VWHEQTPDWEFKDANGN-----TLSSKDALLSR 453
QY 259 AGAEIFGVLVYFGAYKDLQATPDWNSVKGN--YSW--FVDYLDQMRLLSQAQEKRLLDVF 316
DB 454 LKQVIYTVW---GRY-----KGVYANDVNVNEAIDE-----SOGNGFR----- 488
QY 317 DVHWYTPMAMGGGIRITNEVGNDETCKKARMOAPRTLMDPTYK-----EDSWIAQWNS 368
DB 489 RSNWY-----NICGPEVIERAFIWAHEA--DPDAKLFYNDYNTENSOKRQFIYN 535
QY 369 FL-----PLLPRLKQSVDKY--YPGTKLALTE-----YSYGG 398
DB 536 MIKSLKEKGVPIHIGLQCHINDWPSISEINTIKLFSSIPGLETHITELDMSFPQWGS 595
QY 399 END-----ISGGTAMADVLGLGK--NDVYMANVWKLKDCANNVYSAALKLRNVDG 448
DB 596 STSYSTPPRDLILLKQARVKELEDFLKKYNVITNTVFWGLKD-----DYSW 641
QY 449 KSTFGDISVHAQTSDIVNSSVHASVTDASYKELHLVVMNKSMDSAFDAQFLSGETTYG 508
DB 642 LSONFG-----KSDYPLLFDCNGYKSKY-----AF----- 665
QY 509 SGKVMGFDKNSOIKAAPIQTISGNRFTYVPPPLTAYHIVLTAGNDTPVPENPESFALRA 568
DB 666 -----WSLIEPT-----VVPVNS-----TLFAPPAIQVPTPTPTPTVTSATPT 708
QY 569 EAGD-----GKSYLDASSGWGYSVORATYENGPPAAVASNLVETSYDNTVNTGTSYY 624
DB 709 PAPTASPAGGSYWTFS-----YGALKVWYANG-----NMSST-----TNVLPN----- 748

QY 625 KITAKTKTGSASNLKAVPR--APVDGPDORYEAEDGTLKGTVVESSTGSGTGYVTNF 682
DB 749 KIKIE-NVGTTAVDLRSRVKRYWYITDG-----EAAQSV---SVASSINPAYIDRVVVKLG 800
QY 683 HNAGDSLMTWITQATAGLYNLITGYRSPHDDKRTNFSLNGKASGELVLMKTADEFKETS 742
DB 801 ANAG-----GADYVVEVGFKS-----G 817
QY 743 KVLNLNAGANT-----IGFETGKMWYDIDYVKLEPAADRPPHAYTKTLINPNATVEAKALMN 798
DB 818 AGVLAACQSKTEKRLSKQKSSGTYN----- 842
QY 799 YLVDQYGNLMSQEDMPEIDWLQANVKKPAITAALDLIDYSPSRAEHGLSSTETEKATE 858
DB 843 -----QSN-----DYSVRS-----NSYIE 857
QY 859 WDKOGGIVTFAWHNAPKGLIDTQGEWWRGIFYADSTTFDIEYAMNHPESEDKYLIRDI 918
DB 858 NEKVTGYIDDLVLMGREGPR--NAQIKVW-----YANGNLSPTNVLPN-----KIKIENV 906
QY 919 DVIAGQLKLQDAKVPVLPRLHEAEGKWFWMGAKGPEPVKKLYILMHDRLTNVHKLNL 978
DB 907 GTTAVDLRSV---KVR-----YWTIDG-----EATQS 931
QY 979 IWVWNSVADPWYP-----GDEYVDI-----LSFDS 1003
DB 932 VSVASSINPAYIDRVVVKLGANAGGADYVVEGFKSGAGVLAAGQSTKEIRLSIQKSSGS 991
QY 1004 YPOAGDYSPQIAKYEDLVTLGRDKLVCHERKRTYPCGSDGVSPLBELVYMWGFL 1063
DB 992 YNOSNDYSVRS-----NSYT---ENEKVTGYIDV--LVWGKEP 1026
QY 1064 RRQTKOSLEHLKVVYVNHVWITLKLPTNLKTVGITEQPSVPGSFTLNAAGETAKVLSW 1123
DB 1027 SRGTKPAGE-----VTPAPTPTPTPTPTTPA--PTSAPTPTPTPTPTPTPT 1074
QY 1124 TASANAASIEVRSRVENGAFATVADSVYGSSTYDTAVTADTYYYVYVAKNDAGQTVSN 1183
DB 1075 TPTATP-----TPTPTPTPT-----PTPTP 1094
QY 1184 TASAPKADTOOPT-----TGLVLOVRTADTNVNDNLNPHFQILNKGTISVPINELKI 1237
DB 1095 TVTVTP---TPTPTGPTGSGLKVLVYKNETSASTGSRPMPFKIVNGSSSDVLSRVKI 1151
QY 1238 RYYTIDGREGTFCNDYAVLSCSKLNGKLVKMDKAAATGADYILEVSNFNSAGVLSAGS 1297
DB 1152 RYWTYVDGDKPQSCVCDNAQIGASNVTFNFVKLTSGVSADYILEVFGSSGAGQLQPKD 1211
QY 1298 TGGIOTRTHKADWSNYSDDSYKGTQTSFDDHTKATLYHNGVLVWGTE 1347
DB 1212 TGDIOVRFNKNDWSNYNQADWSNMOSMTNYGENAKVTLYVDGVLVWGQE 1261

RESULT 7
A43802
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellu
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Caldocellum saccharolyticum
C:Date: 30-Jan-1993 #sequence_revision 30-Sep-1993 #text_change 10-Jul-1998
C:Accession: A43802
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P
Appl. Environ. Microbiol. 56, 3117-3124, 1990
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophil
A:Reference number: A43802; MUID:91136262
A:Accession: A43802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 1-915 <SAU>
A:Cross-references: EMBL:X13602
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 5.5%; Score 397.5; DB 2; Length 915;
Best Local Similarity 27.3%; Pred. No. 1.3e-14;
Matches 126; Conservative 79; Mismatches 162; Indels 95; Gaps 21;

QY 941 HEAEKGFVWGAAGKPEPVK-----KLYILMHDRLTNVHKLNNLIWVNSV-----A 986
DB 77 HNOTPDWFFKDSNGQRLSKDALLARLKQYI-----YDVVGRYKGVYADVWVNEAIDENQ 131
QY 987 PD-----WYP--GDEYVD---ILSFDSPQA---GDYSPOIAKYED---LVTLGKDK 1027
DB 132 PDSYRRSTWYEICGPEYIEKAFIWAHEADPNKLFYNDYNTSEIKKRDFIYNNVKNLKS 191
QY 1028 KLVCHERKRTYPGSGDEGVSSPLELVRYVMVGFLERRQTKQSLHKLKVVNPNV-ITL 1086
DB 192 GIPIH-----GIGMQCHI--NVNW-----PSVSEIENSILFSSIPGIEIH 230
QY 1087 EKLPNTLTKYITGEQSPVPGSFTLNAAAGETAKVKSWSANAASEYVKRSTVENGAFAT 1146
DB 231 TELDMSLYNGSSSENSTPPQDILQK--QSKYKEIF-----MLKYYKNVYKSVTFWG 282
QY 1147 VASD-----VYGS-----YTDTAVTADTYYYOVVAKNDAGQTVSN-----TASA 1187
DB 283 LKDDYSWLSRFGKNDWPLFFED--YSAPKPAYWVIEA---SGVTTSSPTPTPTPTVV 337
QY 1188 APKADTQPTTG--LVLYQRTADTNVNDNLHNFHQLNKGITISVPINELKIRYYTIDG 1245
DB 338 TP-TPTPTPGGQIKVLYANKETNSTNTIRPWLKVVNSGSSSIDLSRVTIRYWTVDG 396
QY 1246 DRETFNCDYAVLSCSKLNGKLVKMDKAATGADYILEVSNFSDAGVLASGGSTGGTQTRI 1305
DB 397 ERAQSAVSDNAQICASNVTFFKVLSSVSGADYILEIGFKSGAGQLQPGKDTGEIQIRF 456
QY 1306 HKADWSNYESDDSYSGTGTSPDHPKATLYHNGVLVWGTGE 1347
DB 457 NKSDWSNNOGNDWSLQSMTSYGENEKVYATIDGVLVWQGE 498

RESULT 8

S02711
cellulase (EC 3.2.1.4) precursor - Caldocoellum saccharolyticum
N:Alternate names: endo-1,4-beta-glucanase
N:Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C:Species: Caldocoellum saccharolyticum
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: S02711
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A:Title: Nucleotide sequence of a gene from Caldocoellum saccharolyticum encoding for exo-
A:Reference number: S02711; MUID:89098398
A:Accession: S02711
A:Molecule type: DNA
A:Residues: 1-1039 <SAU>
A:Cross-references: EMBL:X13602; NID:q40645; PIDN:CAA31936.1; PID:q40646
C:Genetics:
A:Gene: celB

C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1039/Product: cellulase #status predicted <MAT>
F:72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 5.5%; Score 397; DB 2; Length 1039;
Best Local Similarity 27.5%; Pred. No. 1.7e-14;
Matches 131; Conservative 71; Mismatches 169; Indels 106; Gaps 21;

QY 941 HEAEKGFVWGAAGKPEPVK-----KLYILMHDRLTNVHKLNNLIWVNSV-----A 986
DB 129 HNOTPDWFFKDSNGQRLSKDALLARLKQYI-----YDVVGRYKGVYADVWVNEAIDENQ 183
QY 987 PD-----WYP--GDEYVD---ILSFDSPQA---GDYSPOIAKYED---LVTLGKDK 1027
DB 184 PDSYRRSTWYEICGPEYIEKAFIWAHEADPNKLFYNDYNTSEIKKRDFIYNNVKNLKS 243
QY 1028 KLVCHERKRTYPGSGDEGVSSPLELVRYVMVGFLERRQTKQSLHKLKVVNPNV-IT 1085
DB 244 GIPIH-----GIGMQCHI--NVNW-----PSVSEIENSILFSSIPGIEIH 282
QY 1086 EKLPNTLTKYITGEQSPVPGSFTLNAAAGETAKVKSWSANAASEYVKRSTVENGAFAT 1123
DB 283 TELDMSLYNGSSSENSTPPQDILQK--QSKYKEIF-----MLKYYKNVYKSVTFWG 342
QY 1124 TASANAAS-----YEVKRSTVENGAFATV-ASDYGSSYTDFTAVTADTYYYQVVA 1173
DB 343 LRSFYGKNDWPLFFEDYSAPKA---YWAVIEASGVTTSSPTPTPTPTVT-----VTP 392
QY 1174 KNDAGQTVSNTASAP--KADTQPTTG--LVLYQRTADTNVNDNLHNFHQLNKGITISV 1230
DB 393 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 452
QY 1231 PINELKIRYYTIDGREGTFCNDYAVLSCSKLNGKLVKMDKAATGADYILEVSNFSDAG 1290
DB 453 DLSRVTIRYWTVDGERAQSAVSDNAQICASNVTFFKVLSSVSGADYILEIGFKSGAG 512
QY 1291 VLAGSGSTGGITRIHAKDWSNYESDDSYSGTGTSPDHPKATLYHNGVLVWGTGE 1347
DB 513 QLQPGKDTGEIQIRFNKSDWSNNOGNDWSLQSMTSYGENEKVYATIDGVLVWQGE 569

RESULT 9

T31337
1.4-beta-glucanase (EC 3.2.1.4) - Anaerocellum thermophilum (fragment)
C:Species: Anaerocellum thermophilum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31337
R:Zverlov, V.; Mahr, S.; Riedel, K.; Bronnenmeier, K.
Microbiology 144, 457-465, 1998
A:Title: Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) f
omains.
A:Reference number: Z21003; MUID:98154434
A:Accession: T31337
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1711 <ZVE>
A:Cross-references: EMBL:Z86105; NID:ei071329; PID:e350354; PIDN:CAB06786.1
C:Genetics:
A:Gene: cela
C:Keywords: glycosidase; hydrolase

Query Match 5.5%; Score 391.5; DB 2; Length 1711;
Best Local Similarity 20.8%; Pred. No. 7.8e-14;
Matches 237; Conservative 149; Mismatches 421; Indels 331; Gaps 43;

QY 416 GKNDVYMANVWK-----LKGANNVVS-----AAKLYNRVDGKSST-----FGD 455

DB 22 GKLPNNVNNNRGDSALKDGDNDGLDTGGWFDAGDHVKNFPLPMSTYTMLSWVVEYKD 81

QY 456 ISVHA-QTSDIVNS-----SVHASVTDASYKELHLVVMKMSDMSAFDAQDLSET 505

DB 82 AFVKSGLQLEHLIAQIEWNDYFKVCHPSKVYYIQ-----VGDGSKDHW----- 126

QY 506 TYSGKVGWGFKNSSQIKAAPIQISGNRFTYVPLPPLTAYHVLVIFAGNDTPVENPESA 565

DB 127 -WGAPEVMQMERPSFKVTQSSPGSTV---VTETAASLAASIVLKDRNPRTKAATYLA 181

QY 566 LRAEAGDGKSIY--LDASSGVVGVSVQRATYENGPPAAVASNLVETSYDTT---NVTNCT 620

DB 182 -----KELYFAEVTKSDAGY-----TAANGYNSWSGFDYDELSWAALWLYLATNDS 228

QY	526	APITQISGRNFTVTPPLTAY	-----HIV--LTAGNDTPVE-NPESFALRAE	569
Db	86	LPMS-----YGTMLSWAAEYKDAFVKS	GQLEHILNQIEWNDIFYKCHPSKYVYYQ	139
QY	570	AGDG-----KSIYDASSGVVGSVORATY	ENGPPFAAVASINLVETSYTDTNVTNG	619
Db	140	VGDDGKHAWGPAEVMQMERPS----	EKVTSQSPGSVAVAETAASLAAS-	186
QY	620	TSYXXKITAKTTGTSASNLKAVPRAPVDG	PRDYEAEDGTLKGTIVESSGTGP-SGTGY	678
Db	187	-----IVLKDRNPKAATYLO-----	HAKDLYEFAE-----VTKSDSGYTAANG	226
QY	679	VTFNHNAGDSLFTMTIQAPTAGLYNLT	-----IGYRSPH--D	712
Db	227	YNSWGFYDELSWA-----AVWLYLATND	STVLTKAESYVQWNPKISGSIIDYKWAHCW	282
QY	713	DKRTNFSLNGKASGELVWKTKADFRETSGG	KVLLNAGANTIGFE-----TGWGYD-	763
Db	283	DVH-----NGAA-----LLAKITD-	KDTYKQIESHLDYWTGTGYNGERIKYTPKGLAWLDQ	333
QY	764	-----IDYVKLEPAADRPPHAVTKTLINPN	NATVEAKALMYLDVQYCKNMLSQEDMPEIDW	820
Db	334	WGLRAY-----ATTTA-----FLAFVYS	-----DW	353
QY	821	LOANVCKKPATAAL--DLIDYS-----	PSRAEHLGSTTEKAIEWDKQ	862
Db	354	SGCPTGKETYKKGESOIDYALGSTGRSFV	YGFGTNPXPPRPHRTAHS-----SWADS	407
QY	863	GGIVTFAHWNAPKGLIDTCQEKWMRGFYAD	STTFDIEYAMNHPESEDYKLLIRIDVIA	922
Db	408	OSIPSYYRHH-TLYGALYGGPGSD-----	DSYTDLISNVVNEVACDYN-----A	450
QY	923	GOLKKLQDAKVPVLF--RPLHE-----	AEGKWF-----WW	950
Db	451	GFVGAL--AKMYLLYGGNPIDPFKAIETPT	NDEFFVEAGINASTGPNIEIKAINVNSQSGW	508
QY	951	GAKGPEPVK-KLYILMHDRLTNNVHKLNNLI	-----WYWN5-----VAPDW	989
Db	509	PARATNKLKFRYFVDLSLSELIKAGYSPNQL	TLSYTNQCAKVSFPYVWDSERNYYILVD	568
QY	990	-----YPG--DEYVDILSFD-----	SYPOAGDYSQPIAKYEDLVTLGDKDKLVCHERK	1035
Db	569	TGTLIYPGQDKYKKEVQFRIAAPQNVQNS	INDYS-----FQDI-----	608

RESULT 10
T17120
cellulase (EC 3.2.1.-) precursor, thermoactive - Caldocellum saccharolyticum
C-Species: Caldocellum saccharolyticum
C-Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C-Accession: T17120; A43745
R-Te'o, V.S.; Saul, D.J.; Bergquist, P.L.
Appl. Microbiol. Biotechnol. 43, 291-296, 1995
A-Title: Cella, another gene coding for a multidomain cellulase from the extreme
A-Reference number: Z18698; MUID:95336703
A-Accession: T17120

199	-GDGTFGNMTGEDFGQINVNHNISDVKSVPTEIAHSPVDSGVFLADASANDAACKLY	255
798	NYLYDQYGYKNNLSGQEDMFEIDW-----LOANVYGKKPATAALDLIDYS-PSRAEHG-L	848
258	KYLRLVYGNKLTSGM--MAHVAWNHDEADKIHVLTKGYPAINCYDFTHIAPVQSGNSGWI	315
849	SSTETERAIEWDKOGGIVTFPAHWNAPKGLIDTQCKEW-WRGFYADSTTTTDEIYANNHPE	907
316	NYNDITPVTEWADAGGIVLSLWHFNVPQNETTICADSGSGGINSQTTKASHALVSCT	375
908	SEDYKLLTRIDVIDIAGQKLDQAKVPVFLRPLHEAEKG-----WFWMGAKGP	955
376	WEN-KFEMEQQENVANVILKLDQAGIVALWRPFHEAAGNATLKSGANWGKAWFWMGEDGP	434
956	EVVKKLYLLMHDLRTNVVHKLNNLILWVNSV-----APWYPCQDEVDTILGFSDY	100
435	DVYKOLHHTMFNFTSN-KGTHNLIEWEWTSONYNGSDIYNNDDDDWYPCDAYVDIIGRLY	493
1005	-----PQAGDYSPOIAKY	1017
494	GTTAVOOYSEYSOLKGRY	511

RESULT 12

A26874
cellulase (EC 3.2.1.14) precursor - *Bacillus subtilis* (strain DLG)
N:Alternate names: endo-1,4-beta-glucanase
C:Species: *Bacillus subtilis*
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 15-Oct-1999
C:Accession: A26874; B26874
R:Robson, L.M.; Chambliss, G.H.
J. Bacteriol. 169, 2017-2025, 1987
A:Title: Endo-beta-1-4-glucanase gene of *Bacillus subtilis* DLG.
A:Reference number: A26874; MUID:87194581
A:Accession: A26874
A:Molecule type: DNA
A:Residues: 1-508 <ROB1>
A:Cross-references: GB:M16185; NID:g143007; PIDN:AAA22496.1; PID:g143008
A:Experimental source: strain DLG
A:Accession: B26874
A:Molecule type: protein
A:Residues: 39-53 <ROB2>
A:Experimental source: strain DLG
A:Note: the authors believe Met-1 and Met-2 may be alternate initiators
C:Comment: The low molecular weight of the mature protein suggests carboxy-
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucopolysaccharides
A:Pathway: cellulose degradation
C:Keywords: alternative initiators; extracellular protein; glycosidase; hy-
E:1-38Domain: (or 2-18) signal sequence #status predicted <SIG>

[illegible]

QY 1262 KLNGKLVKMDKAATGADYYLEVSFNSDAGVLASGGSTGGIQTIRHKADWSNWNESDDYSY 1321

Db 409 AOLGCGNVTYKFVTLHRPKQCADTYLELGFKN--GTLAPGASTGNIQLRLHNDNDSNYAQ 466
Qy 1316 SDDYSYKGTOTSFDDHTKATLYHNGVLVWGTE 1347
Db 467 SGDISFFKSNT-FKTTKKITLYDQGLIWGTE 497

Search completed: July 13, 2002, 00:48:58
Job time: 20154 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:59:23 ; Search time 32.61 Seconds
(without alignments)
1602.925 Million cell updates/sec

Title: US-09-784-554B-4
Perfect score: 7165
Sequence: 1 MKAKNSSIWSKRSKWLPPV.....HTKATLYHNGVLVWGTTAN 1350

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1797.5	25.1	700	1 GUNA_PAELA	P29719 paenibacill
2	1454.5	20.3	1331	1 MANB_CALSA	P22533 caldocellum
3	822.5	11.5	1021	1 MANA_RHOMR	P49425 rhodothermu
4	717.5	10.0	571	1 MANB_PIRSP	P55297 piromyces s
5	694	9.7	606	1 MANA_PIRSP	P55296 piromyces s
6	690	9.6	569	1 MANC_PIRSP	P55298 piromyces s
7	413.5	5.8	914	1 GUX2_CLOSR	P50900 clostridium
8	397	5.5	1039	1 GUNB_CALSA	P10474 c endogluc
9	372	5.2	1742	1 GUNA_CALSA	P22534 caldocellum
10	356.5	5.0	499	1 GUN1_BACSU	P07983 bacillus su
11	351	4.9	145	1 YCEA_PAELA	P29718 paenibacill
12	348.5	4.9	499	1 GUN3_BACSU	P23549 bacillus su
13	347.5	4.8	499	1 GUN2_BACSU	P10475 bacillus su
14	339.5	4.7	1853	1 CIPA_CLOTM	P04755 clostridium
15	337	4.7	772	1 GUNP_CLOTM	P01866 clostridium
16	328	4.6	986	1 GUNZ_CLOTR	P23659 clostridium
17	319	4.5	879	1 GUN1_CLOTM	P02934 clostridium
18	279.5	3.9	505	1 GUNV_ERWCA	P47096 erwinia car
19	277	3.9	504	1 GUNW_ERWCA	P59395 erwinia car
20	275.5	3.8	1475	1 APU_THETY	P16950 t amylopull
21	268	3.7	513	1 MANB_BACSM	P16699 bacillus sp
22	259	3.6	1279	1 APU_THESA	P36905 t amylopull
23	254.5	3.6	1861	1 APU_THETU	P38536 t amylopull
24	247	3.4	360	1 MANB_BACSU	P55278 bacillus su
25	247	3.4	419	1 MANA_PSEFL	P49424 pseudomonas
26	226	3.2	1848	1 CBPA_CLOCL	P38058 clostridium
27	216	3.0	1481	1 APU_THET	P18939 t amylopull
28	191.5	2.7	1902	1 P2P_LACPA	P02470 lactobacill
29	189.5	2.6	1953	1 BIGA_SALTY	P25927 salmonella
30	189	2.6	1433	1 SUBF_BACSU	P16397 bacillus su
31	187	2.6	872	1 GUXA_CELFI	P50401 cellulomona
32	179.5	2.5	964	1 CTA2_BACCI	P70873 bacillus ci
33	178.5	2.5	972	1 CTA1_BACCI	P94286 bacillus ci

ALIGNMENTS

```
RESULT 1
GUNA_PAELA
ID GUNA_PAELA STANDARD; PRT; 700 AA.
AC P29719;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase A) (EG-A).
GN CELA
OS Paenibacillus lautus (Bacillus lautus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Paenibacillus.
OX NCBI_TaxID=1401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PL236;
RX MEDLINE=92276330; PubMed=1592807;
RA Hansen C.K., Diderichsen B., Joergensen P.L.;
RT "celA from Bacillus lautus PL236 encodes a novel cellulose-binding
  endo-beta-1,4-glucanase.";
RL J. Bacteriol. 174:3522-3531(1992).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
  linkages in cellulose.
CC -!- PTM: A SHORT FORM (EGA-S) ARISES FROM POSTTRANSLATIONAL
  PROTEOLYSIS OF APPROXIMATELY 150 AA AT THE C-TERMINUS OF EGA-L.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY J (FAMILY 44 OF GLYCOSYL
  HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation at
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76588; AAA22303.1; -.
DR PIR: B41897; B41897.
DR PIR: S27499; S27499.
DR HSSP: Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR Pfam: PF00942; CBD_3; 1.
DR ProDom: PD001947; CBD_3; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 33
FT CHAIN 34 700
FT DOMAIN 548 700 ENDOGLUCANASE A.
FT ACT_SITE 213 213 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 213 213 BY SIMILARITY.
SQ SEQUENCE 700 AA; 76910 MW; 3D5C8CADA53EE0F CRC64;
```

Query Match 25.1%; Score 1797.5; DB 1; Length 700;
Best Local Similarity 52.0%; Pred. No. 8.6e-91;
Matches 368; Conservative 86; Mismatches 183; Indels 71; Gaps 13;

Qy	11	SKRSKWLPPVWACTIIIVGGALPPTVVVGHGTAKTVIKVDTSKDRKPISPYIYGTNOELA	70
Db	3	TRQKRLEFVSAALAVSTMTVPMPASVNA--AASDVFTTINTQSERAASIPNIYGTNODLS	61
Qy	71	GDENLTRRLLGGNRMTGYNWNNNNNAGSDWMOSSDSYLCDNAGLTAKACEKPGAVATSF	130
Db	62	GTEWSSRRLLGGNKLCTGYNWNNNASSAGRWMLHYSDFLCGNGGVDPDTCCKPGAVVTAF	121
Qy	131	HQQLSKOGTYSLVTLPMAGVYVAKDNGSGVQSEKAPSARWNVVNAKNAFPQLOPLDKON	190
Db	122	HDKSLENGAYSIVTLQAGVYVRDKNCPVDESETAPSPWDKVEFAKNAFSPLOPHLNDG	181
Qy	191	QVYADEVFNLVKYGVASTKTVGVKYSLDNEPALMWSHTRPHGEBKVAKELVDRSVSL	250
Db	182	QVYMDVEEVLNARYGNASTSTGIKAYSLDNEPALMSETHPRHPQLQAAELVAKSIDL	241
Qy	251	SKAAKAVDACAETFGPVLYGFGAYKDOTAPDMNSVKGNYSWFVDXYLDQMLRSSOAEK	310
Db	242	SKAVKNVDPAEIEFGALYFGGAYLSLQDAPGWSLGGNYSWFIDYLDQMLNAHTQNGK	301
Qy	311	RLLDVFDVHWYPEAMGGGIRIT--NEYGNDETKARKMOAPRTLWDPTYKEDSWIAQWNSF	369
Db	302	RLLDVLDVHWYPEAQGGORIVFEGGAGNIDTQKARVQAPRSLWDPAQEDSWIGTFWSSY	361
Qy	370	LPILLRLKQSDVKYYGPTKALTEYSYGGENDISGGIAMADVILGLKNDVYMANVWKJK	429
Db	362	LPILPKQSSIQIYYPTGLKATIESYGGDNHISGGIATADALGIFGKYGVYAANYWQTE	421
Qy	430	DGANNVYSAAYKLYRNYDCKSPFGDISVHAQTSIDTVNSSVHASVTDASYKELHLVVWVK	489
Db	422	DN--TDITYSAAYKLYRNYDGNKSGFGIKVDAATSDTENSVSVASTDENSELHLIVLNK	480
Qy	490	SMDSAFDAQFDLSGTTYGSGKVMGFDKNSSQITKEAAPITQISGRNFTYVPLPTAYHTV	549
Db	481	NFDDPINATFQLSGDKTYTSGRWGFDQSGDITEQAAITNNINNOFTYLPPLSAYHLV	540
Qy	550	LTAGNDTPVENPESPALRAEADGKSIYLDASGVVGVYQVRATYENGFFAAVASNLVET	609
Db	541	LKADSTEV-----NSDLV-----VOYKDGDRNATDQIKP	572
Qy	610	SYTDITVNTN-GTS-----YYYKITAKTKTGTSSASNLVKAVPRAPVDGPD---RYE	655
Db	573	HF---NIQNGKTPVDLSLTLRYFP-----TDSSAAMN--GWIDWAKLGSNTQISFG	622
Qy	656	AEDGTLKGTWVE---SSGTGFGSGTVV-----TFNHNAGD	687
Db	623	NHNGADSDTYAELGFFSSGAGSIAGGQSGEIQLRMSKADWSNFNEAD	670
RESULT	2		
MANB_CALSA			
ID	MANB_CALSA	STANDARD;	PRT; 1331 AA.
AC	P22533;		
DT	01-AUG-1991	(Rel. 19, Created)	
DT	01-JUL-1993	(Rel. 26, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Beta-mannosidase/endorucanase A precursor	[includes: Mannan endo-1,4-	
DE	Beta-mannosidase A (EC 3.2.1.78)	(Beta-mannanase) (Endo-1,4-	
DE	mannanase); Endo-1,4-beta-glucanase	(EC 3.2.1.4) (Cellulase)]	
GN	MANA.		
OS	Caldocellum saccharolyticum	(Caldicellulosiruptor saccharolyticus).	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Thermoanaerobacter group; Caldicellulosiruptor.		
ON	NCBI_TaxID=44001;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=93J19139; PubMed=1476429;		
RA	Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;		
RT	"The beta-mannanase from 'Caldocellum saccharolyticum' is part of a		
RT	multidomain enzyme."		
RL	Appl. Environ. Microbiol. 58:3864-3867(1992).		
RL	[2]		
RP	SEQUENCE OF 1-346 FROM N.A.		

QY 190 NOVYADEFVFLVKKYGVASTTKYGVYSIDNEPALWSHTPHIHCCKYGAELVDRSVS 249
 DB 932 NFVYMEFINLYLINKYGMASSTPTGIKGYITLDNEPDLWASTHPRIHPNKVYCKELIEKSV 991
 QY 250 LSKAAKAVDAGAEIFGFLVLYGFCAYKDLQATPDWNSVKGNYSWFVDYLDQMLRSSQAE 309
 DB 992 LAKVITLPSAEVFCYASYGMYYSLODAPDNQVKGHEHWFISWYLEQMKKASDSFG 1051
 QY 310 KRLDVFVHVHTPEAMGGGIRITNEVGNDETCK---ARMOAPTTLWDPTYK----- 357
 DB 1052 KRLDVLDLHWYPEARGGNIRVCFDENDTSKEVVIARMOAPTTLWDPTYKTSVKGOITA.1111
 QY 358 -EDSWAONNSFELPLLRKOSVDKYPPTKALATEXSYGENDISGGIADVLGILG.416
 DB 1112 GENSINQWFSYDPLPIPNVKADIEKYPTKLAISEPDYGNHNSIGGIALADVLGIFG 1171
 QY 417 KNDVYMYNTWKLDGANNVVSAAKLYRNYDGKSTFGDISVHAOTSDIVNSVHASVTD 476
 DB 1172 KYGVNFAARW---GDSGSYAAANYYLNYDCGSKYGNVTNSANTSDVENNPPVVASING 1228
 QY 477 ASYKELHLVYMKSMDSAFDAFDLSGETTYGSKGVGDKNKSQIKEAPTQISGNRF 536
 DB 1229 QDSELEHILINRNVDKQLQVKINITSPTKYTKAEIYGFDSNSPYKMGNDINDIESNVF 1288
 QY 537 TYTVP 541
 DB 1289 TLEVP 1293

RESULT 3
 MANA_RHOMR STANDARD; PRT; 1021 AA.
 AC P49425;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase (EC 3.2.1.78).
 GN MANA.
 OS Rhodothermus marinus (Rhodothermus obamensis).
 OC Bacteria; CFb group; Rhodothermus group; Rhodothermus.
 OX NCBI_TaxID=29549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 43812;
 RA Pollitz O., Krah M., Borriess R.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucmannans.
 CC -!- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X90947; CAA62442.1;
 DR InterPro; IPR000805; Glyco.hydro_26.
 DR Pfam; PF02156; Glyco.hydro.26; 1.
 DR PRINTS; PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 1021 AA; 115791 MW; A8F7B63109E0F715 CRC64;

Query Match 11.5%; Score 822.5; DB 1; Length 1021;
 Best Local Similarity 39.1%; Pred. No. 2.3e-37;
 Matches 194; Conservative 81; Mismatches 164; Indels 57; Gaps 14;
 QY 648 VDGPDREYEDGTLKGTVESSGTGSGTYVTNPHNAGDSLTMWTQAPTAGLYNLITIG- 706

DB 476 VAGEIRLEAEDGCELLGVAVDSTLTCYSGRGVYVTCFADAPEDSVRFSEAP-RGYRVVFGV 534
 QY 707 -----YRSPHDKRTNFSLANGKASGELVLMKTADPKETSGKVLNAGANTIGFETG 758
 DB 535 SFSSRFASIALRVDDHWQTSGLIKRGG-----FFEASIGEILWDEGAHTMAFOLM 585
 QY 759 WGWYDIDYVVKLEPADRPPHAYTKTLINPNATVEAKALMNYLDVOYGNMLSQGE----- 813
 DB 586 NG--ALDYVRLEPVSYGPPARPPAQLSDSQATASAQALFAPLJSEYGRHLIAGQQNPYR 643
 QY 814 -DMPEIDWLOANVGKPAIALDLIDYSPSRAEHLGSSTET-EKAIEDWKOGGIVTFAWH 871
 DB 644 RFDAINVVRNVNTGKEPALVSFDLIDYSPREAHGVVHYQTPEDWIAWAGRDGIVSLMWH 703
 QY 872 WNAPKGLIDTOGKE--WNRGFYADSTTFDIEYAMNHPESEDYKLLIRIDIVIAQLKKLQ 929
 DB 704 WNAPTDLIEDPSQDCYWMYGFYTRCTTFDVAALADTSSERYRLLLRDIDVIAAQLQKFQ 763
 QY 930 DAKVPVLFRLPHEAGKFWWGAKEGPEVKLYILMHDRLTNVHKLNLIWVNSV--AP 987
 DB 764 QADIPVLMWRPHEAAGGFWWGAKEGPEFKQLMRLLYERLVHHLHGLNLIWVYTHEPGAA 823
 QY 988 DWYPCDEVVDILSPDSYPOAGDYSQIAKYEDLVTLGKDKLVCHERKRTYPPGSGDEGV 1047
 DB 824 EWYFGDAYVDVIGRDVADDPD-ALMRSDWNEQLTLEGGKLVALTETGILP----- 874
 QY 1048 SSPLLELVY-YMWGFLERROTQSL-----EHLKKVYNHPNVITPLEKPLNLTGYGTEQ 1101
 DB 875 --DVEITDYGIWNSWFSIWTDPFLRDVDPDLRTVYHSERVLTRDELDP-DWRSY----- 926
 QY 1102 PSVPGSFTLNAAGETA 1117
 DB 927 --VLHATTVPAGDLA 940

RESULT 4
 MANB_PIRSP STANDARD; PRT; 571 AA.
 AC P55297;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78) (Beta-
 DE mannanase B) (1,4-beta-D-mannan mannanohydrolase B).
 GN MANB.
 OS Piromyces sp.
 OS Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
 OC Neocallimasticeae; Piromyces.
 OX NCBI_TaxID=45796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96313314; PubMed=9768520;
 RA Millward-Sadler S.J., Hall J., Black G.W., Hazlewood G.P.,
 RA Gilbert H.J.;
 RT "Evidence that the Piromyces gene family encoding endo-1,4-mannanases
 RT arose through gene duplication".
 RL FEMS Microbiol. Lett. 141:183-188 (1996).
 CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucmannans.
 CC -!- SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANA, MANC
 CC AND XNA; AND TO THOSE OF N.PATRICIARUM XNA.
 CC -!- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; X97408; CAA66061.1; --
 DR InterPro; IPR002883; CBD_5;
 DR InterPro; IPR000805; Glyco_hydro_26.
 DR Pfam; PF02013; CBD_5; 2.
 DR Pfam; PF02156; Glyco_hydro_26; 1.
 DR PRINTS; PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase; Signal; Multigene family; Repeat.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 571
 FT DOMAIN 490 571
 FT REPEAT 490 571
 FT REPEAT 533 571
 FT DOMAIN 475 479
 FT DOMAIN 482 488
 FT POLY-ASN.
 SQ SEQUENCE 571 AA; 64397 MW; 813E44581FAA9DAA CRC64;

Query Match 10.0%; Score 717.5; DB 1; Length 571;
 Best Local Similarity 32.7%; Pred. No. 5.3e-32;
 Matches 183; Conservative 87; Mismatches 207; Indels 83; Gaps 17;

Qy 654 YEADGTLKG-TVVESSGTGFGTGYVTFNHNAGDSLTMTIQAPTAGLYNLTIGYRSPHD 712
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 24 YEAEKGLDGYTVYKSDLTGTFSGTGYVGRFENPGNSVTVEVPTGMYDLTIVYCANMG 83
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 713 DKRTNLSLNGKASGELVLWKTDKFTSGGKVLNAGANTIGETGHWYDIDYVKLEPA 772
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 84 QKINSLVNGOSAGDITFTENTKREFDNLVGAIVLNKGNKTIGLVSSGWMVDAFVINDA 143
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 773 ADPPHAYTK---TLINPNATVEAKALMNYLDVQYGNKMLSGO-----EDMPEIDWL 821
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 144 PNAAKDVTSLKPLNTPINPAKPAKLYDLKTYNGKRIISLGGVGAAGAGDGGELQRI 203
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 822 QANVGKPAIAALDIDYSPSRAEHLGSLSTETEKAEWDK--QG-GIVTFPAWHNAPKGL 878
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 204 QKATGKLPVANNMDFEFESNDCTWRPQNPDPITEMAINMWWKKGKIMSQAQWHN----- 258
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 879 IDTGKEMWGRGFYADSTTDEIYAMNHPSEDEYKLLIRDIDVLAGLKLQDAKVPVLF 938
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 259 --IAGTKGDFAFYKOTFTSIDNAVTEGTWE-YEKIKIDIDRVAGHIKKLQAVNMP 315
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 939 PLHAEGKFWMGAKGPEPKYILMHDRLTNVHKLNNLIWVNSVAPDWYPCGEYVDI 998
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 316 PLHENDGDFWFG--NNPKSCAKLWKILYERVMYHGLNLIWLNKNDANTPVD-YIDI 373
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 999 LSFDSYPOAGDYSQIAKYEDLVTLGKDKKLVCHEKRRYTPGPGSDGVSSPDELRYMV 1058
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 374 IGVDIY--ANDHGPQTAYNTHEFDYGGKKWVLSSENGRIPD-----IQOCV 418
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1059 -----WGFLERRRQKSL-----HLKKVYNHPNVITLEKLPNT-NLKYI----- 1096
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 419 DONAWNGYFQTNWSEFLLQDSYHTDAQLKEYFTHTKVTMMNDLPSFNVNSYEYQSGNN 478
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1097 -----GITRQPSVPGSFT-----LNAAGETAKVKLSWTASANAASIEYVKRST 1138
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 479 NSSNNNNNSSECFSLPYGPPCKGNTVYVYTDNDGDWGVNENWECIGNSSAVVCWSE 538
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1139 VENGAFATVADSVYGSYTD 1158
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 539 ALGYPCCVSSSDVY---YTD 555
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 RESULT 5
 MANA_PIRSP STANDARD; PRT; 606 AA.
 AC P55296;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase A precursor (EC 3.2.1.78) (Beta-mannanase A) (1,4-beta-D-mannan mannanohydrolase A).
 GN MANA.
 OS Pironyces sp.

OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Pironyces.
 OX NCBI_TaxID=45796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96094325; PubMed=7493964;
 RX Fanutti C., Panyi T., Black G.W., Hazlewood G.P., Gilbert H.J.;
 RT "The conserved noncatalytic 40-residue sequence in cellulases and
 RT hemocellulases from anaerobic fungi functions as a protein docking
 RT domain.";
 RL J. Biol. Chem. 270:29314-29322(1995).
 CC !- FUNCTION: HYDROLYSES 1,4-BETA LINKED POLYSACCHARIDE BACKBONES OF
 CC MANNANS. ONE OF THE MAJOR HEMICELLULOSE COMPONENTS IN HARDWOODS
 CC AND SOFTWOODS. SHOWS VERY HIGH ACTIVITY AGAINST MANNOHEXAOSE BUT
 CC NOT AGAINST MANOPENTAPOSE AND SMALLER MANNOOLIGOSACCHARIDES. THE
 CC MAJOR PRODUCTS RELEASED FROM MANNOOLIGOSACCHARIDE HYDROLYSIS ARE
 CC MANNOSE AND MANNOBIOSE. THE REITERATED 40 AA DOMAIN IS INVOLVED
 CC IN BINDING THE CELLULOSE-HEMICELLULOSE COMPLEX.
 CC !- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic
 CC linkages in mannans, galactomannans, glucomannans, and
 CC galactoglucomannans.
 CC !- DOMAIN: CONSISTS OF A CATALYTIC N-TERMINAL DOMAIN LINKED TO A
 CC REITERATED NON-CATALYTIC C-TERMINAL DOMAIN.
 CC !- SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANB, MANC
 CC AND XYNA; AND TO THOSE OF N.PATRICIARUM XYNA.
 CC !- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X91857; CAA62968.1; --
 DR InterPro; IPR002883; CBD_5.
 DR InterPro; IPR000805; Glyco_hydro_26.
 DR Pfam; PF02013; CBD_5; 3.
 DR Pfam; PF02156; Glyco_hydro_26; 1.
 DR PRINTS; PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase; Signal; Multigene family; Repeat.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 606
 FT DOMAIN 20 471
 FT DOMAIN 472 489
 FT DOMAIN 490 606
 FT REPEAT 490 528
 FT REPEAT 529 567
 FT REPEAT 568 606
 FT SEQUENCE 606 AA; 68055 MW; 79AAFEFA2725D86 CRC64;
 SQ

Query Match 9.7%; Score 694; DB 1; Length 606;
 Best Local Similarity 32.4%; Pred. No. 1.1e-30;
 Matches 179; Conservative 91; Mismatches 216; Indels 66; Gaps 18;

Qy 654 YEADGTLKGTVVESSGTGFGTGYVTFNHNAGDSLTMTIQAPTAGLYNLTIGYRSPHD 713
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 24 YEADGKLGITVFKELSGFSGKGYVGRFENPGNSVTVDAPATGMVDSLIIYCANMG 83
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 714 KRTNPSLNGKASGELVLWKTDKFTSGGKVLNAGANTIGETGHWYDIDYVKLEPA 773
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 84 KINSLTVNDQSVGDITFTENTKFTKDVGAIVLNKGNKTIGLVSSGWMVDAFVINDAP 143
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 774 DRPPHAYTK---TLINPNATVEAKALMNYLDVQYGNKMLSGO-----EDMPEIDWLQ 822
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 144 NAAKDVSSKLNPTLVNPAKPAKLYDLKTYNGKRIISLGGVGAAGAGDGGELQRIQ 203
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 823 ANVGKKPAIAALDIDYSPSRAEHLGSLSTETEKAEWDK---GGIVTFPAWHNAPKGLI 879
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 204 KATGKLPVANNMDFEFESNDCTWRPQNPDPITEMAINMWWKKGKIMSQAQWHN----- 257
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 879 IDTGKEMWGRGFYADSTTDEIYAMNHPSEDEYKLLIRDIDVLAGLKLQDAKVPVLF 938
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 259 --IAGTKGDFAFYKOTFTSIDNAVTEGTWE-YEKIKIDIDRVAGHIKKLQAVNMP 315
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 939 PLHAEGKFWMGAKGPEPKYILMHDRLTNVHKLNNLIWVNSVAPDWYPCGEYVDI 998
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 316 PLHENDGDFWFG--NNPKSCAKLWKILYERVMYHGLNLIWLNKNDANTPVD-YIDI 373
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 999 LSFDSYPOAGDYSQIAKYEDLVTLGKDKKLVCHEKRRYTPGPGSDGVSSPDELRYMV 1058
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 374 IGVDIY--ANDHGPQTAYNTHEFDYGGKKWVLSSENGRIPD-----IQOCV 418
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1059 -----WGFLERRRQKSL-----HLKKVYNHPNVITLEKLPNT-NLKYI----- 1096
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 419 DONAWNGYFQTNWSEFLLQDSYHTDAQLKEYFTHTKVTMMNDLPSFNVNSYEYQSGNN 478
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1097 -----GITRQPSVPGSFT-----LNAAGETAKVKLSWTASANAASIEYVKRST 1138
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 479 NSSNNNNNSSECFSLPYGPPCKGNTVYVYTDNDGDWGVNENWECIGNSSAVVCWSE 538
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Qy 1139 VENGAFATVADSVYGSYTD 1158
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 Db 539 ALGYPCCVSSSDVY---YTD 555
 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
 RESULT 5
 MANA_PIRSP STANDARD; PRT; 606 AA.
 AC P55296;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase A precursor (EC 3.2.1.78) (Beta-mannanase A) (1,4-beta-D-mannan mannanohydrolase A).
 GN MANA.
 OS Pironyces sp.

QY 880 DTQGEKWRGFFVADSTTFDIEVAMNHPESEDKLLIRDDIVAGOLKKLODAKVPVLPFR 939
 DB 258 -TAGTGDFAFYKDTTFNENAVTEGWE-YEKIIDIDRVSGHKKLQAVNMPPLIWRP 315
 QY 940 LHEAGKWFWMGAKGPKVKKLYILMHDLRTNVHKLNNLIWVNSVAPDWYPGDEYVDIL⁴999
 DB 316 LHENNGDNFWWG-NPKKACAKLWKILYERVMYHGLNLIWLNNGNDANTPVD-YIDI 373
 QY 1000 SFDSPQAGDVSPTAKYEDLVTL-GDKDKLVCHERKRTYPCGSDGVSPLRLRYMV 1058
 DB 374 GVDIY--ANDHGPOTTAYNTDFYGGKKMVLVSENGRI---PDIOQCVDOD-----VW 422
 QY 1059 WGFLERRQTKOSLE-----HLKKVYNHPNVITLEKLPNT-NLKYTGITEQPSVPGSFTL 1110
 DB 423 WGFQWTFNSEFILDQSYHTDAQLKEFYFNHKTVMMDLPSFVNSYNGDSSSSNGSES 482
 QY 1111 NAAGETAKYKLSWTASAN-----AASYEVKRTSTVENGAFATVAD-----VYSSGYTD----- 1158
 DB 483 NS--NTGNSDECSINLGYPCIGDYVY--TTDENGDMGVENNEWCGIVHKSCWSEPLGY 538
 QY 1159 -----TAVTAD 1164
 DB 539 PCVGNVTVISAD 550

RESULT 6
 MANC_PIRSP STANDARD; PRT; 569 AA.
 AC P55298;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Mannan endo-1,4-beta-mannosidase C precursor (EC 3.2.1.78) (Beta-mannanase C) (1,4-beta-D-mannan mannanohydrolase C).
 GN MANC.
 OS Piromyces sp.
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimastixales;
 OC Neocallimastixaceae; Piromyces.
 OX NCBI_TaxID=45796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96313314; PubMed=8768520;
 RA Millward-Sadler S.J., Hall J., Black G.W., Hazlewood G.P.,
 RA Gilbert H.J.;
 RT "Evidence that the Piromyces gene family encoding endo-1,4-mannanases arose through gene duplication."
 RL FEMS Microbiol. Lett. 141:183-188(1996).
 CC -|- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.
 CC -|- SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANA, MANC AND XYNA; AND TO THOSE OF N.PATRICIARUM XYNA.
 CC -|- SIMILARITY: BELONGS TO FAMILY 26 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X97520; CAAG6134.1;
 DR InterPro: IPR002883; CBD_5.
 DR InterPro: IPR000805; Glyco_hydro_26.
 DR Pfam: PF02013; CBD_5; 2.
 DR Pfam: PF02156; Glyco_hydro_26; 1.
 DR PRINTS: PR00739; GLHYDRASE26.
 KW Hydrolase; Glycosidase; Signal; Multigene family; Repeat.
 FT SIGNAL 1 18.
 FT CHAIN 19 569
 FT DOMAIN 490 569
 FT REPEAT 488 525

FT REPEAT 531 569 2.
 FT DOMAIN 473 477 POLY-ASN.
 FT DOMAIN 480 486 POLY-ASN.
 SQ SEQUENCE 569 AA; 64115 MW; 19277764E18328B5 CRC64;

Query Match 9.6%; Score 690; DB 1; Length 569;
 Best Local Similarity 32.0%; Pred. No. 1.7e-30;
 Matches 179; Conservative 85; Mismatches 212; Indels 84; Gaps 17;

QY 654 YEAEDETKTGVVSESGTGFSGTYVTFNHNAGDSLTMTIQAPTAGLYNLTIGYRSPHDD 713
 DB 23 YEANGKLDGVSKVYNELOGYSGTGYGVRPESAGNSVTVTVEVSQTGYDMSIIYCGNMGQ 82
 QY 714 KRTNFSLINGKASGELYLWKTADFKETSGGKVLINAGANTIGFTGKGWYDIDYVKLEP-- 771
 DB 83 KINSKLTNGKNGSDITFPENSPFEELNIGAVYLSNGENTVSLVASMGWIWIDALVYNDTP 142
 QY 772 --AADRPHPAVTTLINPNATVEAKALMNYLVQYGNMLSGQ-----EDMPEIDWL 821
 DB 143 NVAKDVSPH-INPTLVNPKAIPAAKKLYDPLRSNYGKRILSGQVGGAGQAGDEGEIQR 201
 QY 822 QANVGKPAIAALDLIDYSPSRAEHLSTETETAKIEWDK--QG-GIVTFANHNAPKGL 878
 DB 202 QKATGKLPAVNMDFIPESNDCTWRPENPDITEMAINNMWKYQGGKIMAAQWHWN----- 256
 QY 879 IDTGKEMWRGFFVADSTTFDIEVAMNHPESEDKLLIRDDIVAGOLKKLODAKVPVLPFR 938
 DB 257 --TAGTKGDFAFYKDKDTTFSIDNAVTEGWE-YEKIIDIDRVSGHKKLQAVNMPPLIWR 313
 QY 939 PLHEAGKWFWMGAKGPKVKKLYILMHDLRTNVHKLNNLIWVNSVAPDWYPGDEYVDI 998
 DB 314 PLHENDGDWFWG-NPKKACAKLWKILYERVMYHGLNLIWLNNGNDANTPVD-YIDI 371
 QY 999 LSFDSPQAGDYSQIAKYEDLVTLGKDKLVCHERKRTYPCGSDGVSPLRLRYMV 1058
 DB 372 IGVDIY--ANDHGPOTTAYNTDFYGGKKMVLVSENGRIPD-----IQOCV 416
 QY 1059 -----WGFLERRQTKOSLE-----HLKKVYNHPNVITLEKLPNT-NLKYT----- 1096
 DB 417 DQNAWGYFOTWNSSEFILDQSYHTDAQLKEFYFNHKTVMMDLPSFVNSYNGDSSSSNG 476
 QY 1097 -----GITEQPSVPGSFT-----LNAAGETAKYKLSWTASANAASYEVKRSR 1138
 DB 477 NSSNNNNNSSECSFIPGLGYPCKGNTVYTDNDGDMGVENNEWCGIGNSSAVVCWSE 536
 QY 1139 VENGAFATVASDVGSSYTD 1158
 DB 537 ALGYPCCVSSSDVY---YTD 553

RESULT 7
 GUX2_CLOSR STANDARD; PRT; 914 AA.
 AC P50900;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II) (1,4-beta-cellobiohydrolase II) (Avicelase II).
 GN CELV.
 OS Clostridium stercorarium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 11754;
 RA Bronnenmeier K., Kundt K., Riedel K., Schwarz W.H.,
 RA Staudenbauer W.L.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION.

QY	1121	LSWTASANAASYEVKSTVENGAFATVADSVYSSYTDATVADTYTQ-VVAKNDAGQ 1179	5.0%; Score 356.5; DB 1; Length 499; Best Local Similarity 35.5%; Pred. No. 2.2e-12; Matches 81; Conservative 36; Mismatches 94; Indels 17; Gaps 7;
DB	286	ISW-VVNWLSKDKESSALAPGASKT-----GGPLTDLTASGTFVRENIRGTDKSTK 337	
QY	1180	TVSNTASAAKADTQPTTGLVLYQYATDNTVNDHNLPHFQILNKGTISVPINELKIRY 1239	
DB	338	DVPET-----PAQDNPTQKGVSVQYKAGDGRVNSQIRPOLHIKNNGNATVLDKDVARY 393	
QY	1240	YYTIDGDEQTFNCDAVYLVSCSKLNGKLVKMDKAAGADYYLEVSPNSDAGVYLVSGSTG 1299	
DB	394	WYVW-KNGQNFCDVQAQMGCGNLTKHFTVTLHPRKQADFTYELGFKT--GTLSPGASTG 450	
QY	1300	GIOTRIHKADWSYNESDDYSYKGTOTSFDDHTKATLYHNGVLVWGTPE 1347	
DB	451	NIQLRLHNDMSYVAQSGDYSFQSNY-FKTTKKIYLIHQGLKWTGE 497	
RESULT 11			
ID	YCEA_PAELA	STANDARD; PRT; 145 AA.	
AC	P29718;		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-APR-1993 (Rel. 25, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein in CELA 5' region (Fragment).		
OS	Paenibacillus lautus (Bacillus lautus).		
OS	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Paenibacillus.		
OC	NCBI_TaxID=1401;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PL236;		
RC	MEDLINE=92276330; PubMed=1592807;		
RA	Hansen C.K., Joergensen P.L., Diderichsen B.;		
RT	"Cella from Bacillus lautus PL236 encodes a novel cellulose-binding		
RT	endo-beta-1,4-glucanase";		
RL	J. Bacteriol. 174:3522-3531(1992).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
EMBL; M76588; AAA22302.1;			
DR	PIR; A41897; A41897.		
DR	PIR; S27498; S27498.		
DR	HSSP; Q06851; INEC.		
DR	InterPro; IPR001956; CBD_3.		
DR	Pfam; PF00942; CBD_3; 1.		
DR	ProDom; PD001947; CBD_3; 1.		
KW	Hypothetical protein.		
FT	NON_TER 1		
FT	SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AEB CRC64;		
SQ			
Query Match	4.9%; Score 351; DB 1; Length 145;		
Best Local Similarity	47.6%; Pred. No. 7.2e-13;		
Matches	69; Conservative 26; Mismatches 48; Indels 2; Gaps 1;		
QY	1202	LQYRTADTNHNLPHFQILNKGTISVPINELKIRYVYITDGDREQTFNCDAVYLVSCS 1261	
DB	1	LQYRTADTNADNQIRPSFNKNGTSADVLSLTFLKIRYFTDKGSAVNGVDWAQLGGS 60	
QY	1262	KLNGKLVKMDKAATGADYYLEVSNFSDAGVYLVSGSTGGTQTRHKAADSNYNESDDYS 1321	

	Query Match	5.0%; Score 356.5; DB 1; Length 499;
	Best Local Similarity	35.5%; Pred. No. 2.2e-12;
	Matches	81; Conservative 36; Mismatches 94; Indels 17; Gaps 7;
QY	1121	LSWTASANAASYEVKSTVENGAFATVASDYGSSYTDTAVTADTTYIQ-VVAKNDAQG 1179
DB	286	ISW-VWNWNSDKOESSALAPGASKT-----GGPLTDLTASGTFFVRENIRGTDKSTK 337
QY	1180	TVSNTASAAPKADPTQPTGLVLQYRTADTVNNDHNLPHFQLNKGTISVPINELKIRY 1239
DB	338	DVPET----PAQDNPTQEGSVQYKAGDGRVNSNOIRPOLHIKNNGNATVDLKDVTARY 393
QY	1240	YYTIDGDEQTFNCDAVILSCSKINGKLVMKAAAGADYYLEVSPNSDAGVLAGSGTG 1299
DB	394	WYNVK-NKGONFCDDAQAOMCGGNLTHKFVTLHRPKOGADFYLELGFKT--GTLSPGASTG 450
QY	1300	GICTRIHKADWSNYNESDDYSYRGKTOTSFDDHTKATLYHNGVLVWGTE 1347
DB	451	NIOQLRHNDMSNYAQSGDYSSFQSNT-FKTTKKIILYHQGKLWGTG 497
RESULT 11		
ID	YCEA_PAELA	STANDARD; PRT; 145 AA.
AC	P29718;	
DT	01-APR-1993 (Rel. 25, Created)	
DT	01-APR-1993 (Rel. 25, Last sequence update)	
DE	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Hypothetical protein in CELA 5' region (Fragment).	
OS	Paenibacillus lautus (Bacillus lautus).	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Paenibacillus.	
OX	NCBI_TaxID=1401;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=PL236;	
RA	MEDLINE=92276330; PubMed=1592807;	
RT	Hansen C.K., Joergensen P.L., Diderichsen B.;	
RT	"Cell from Bacillus lautus PL236 encodes a novel cellulose-binding	
RL	endo-beta-1,4-glucanase";	
CC	J. Bacteriol. 174:3522-3531(1992).	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M76588; AAA22302.1;	
DR	PIR; A41897; A41897.	
DR	PIR; S27498; S27498.	
DR	HSSP; Q06851; INEC.	
DR	InterPro; IPR001956; CBD_3.	
DR	Fam; PF00942; CBD_3; 1.	
DR	ProDom; PD001947; CBD_3; 1.	
KW	Hypothetical protein.	
FT	NON_TER 1	
SQ	SEQUENCE 145 AA; 15782 MW; 9514E3A71B106AE6 CRC64;	

Query Match	4.9%; Score 351; DB 1; Length 145;	
Best Local Similarity	47.6%; Pred. No. 7.2e-13;	
Matches	69; Conservative 26; Mismatches 48; Indels 2; Gaps 1;	
QY	1202	LQYRTADTNVNDHNLPHFOILNKGNTISVPINELKIRYTYITDGREQTFNCDAVILSCS 1261
DB	1	LQYRTADTNVNDHNLPHFOILNKGNTISVPINELKIRYTYITDGREQTFNCDAVILSCS 60
Y	1262	KLNGKLVKMKAATGADYYLEVSNFSNDSAGVLAGSGGTGGTIQTRIHKADWSNYNESDDYS 1321

```

QY 852 ETEKAIEWDKOGGI-VTFAMHWNAPKGLID---TQGEWHRGFPYADSTTFDIEYAMNHPE 907
Db 111 KMKEAVEAARELGIVYIIDHII-----LNDGNPNONKEKAEFFKSSSL---YG-NTPN 161
QY 908 SEDYKLLIRIDIVTAGQLKKLQDAKVPVLPRLPHEAEGKFWGAKGPEVPVKLYILMHD 967
Db 162 -----VIVEIANEPNGDYNWRKDIKPYAEVVISIRKN 194
QY 968 RLTVNHLNLIWV---NSVAPDWPYGDXYDILSFDSPQAGDYSQIAKYEDLVTLGK 1025
Db 195 DPNIIIVGTGWSQDVNDAAADDQLKDANVMDALHF---YAGTHG-QFLURDRKANVALSK 249
QY 1026 DKLVLCHERKTYPGSDGCVSPLVRYVMWG-----FLERQTKQSLEHLK 1075
Db 250 GAPIFVTE-----WGTSASGNGGVFLD-----QSREWLK 279
QY 1076 KVINHPNITILEKPTNLTKYITEQPSVPGSFTLNAAGTAKVKLSWTASANAASVEVK 1135
Db 280 YLDSK---TISWYNWLSKQESSALKPG-----ASKTG---GWRLSLDSAS----- 321
QY 1136 RSTVENGAFATVADSYGSSYTFDFTAVTADTYTYVYVAKNDACOTVNTASAAPKADTQ 1195
Db 322 -----GTF--VRENILG-----TKDST-----KDIPET-----PAKDKPT 349
QY 1196 PTTGLVLQYRTADTNVNDHNLPHFQILNKGITISVPINELKIRYYITIDGREQTFNCY 1255
Db 350 QENGISVOYRAGDSGMSNRQIRPOLQIKNGNTVVDLKDVTARYWNAK-NKQNVDCDY 408
QY 1256 AVLSCKSLNGKLVKMDKAATGADYILEVSNFSDAGVLASGGTGGIQTIRHKADWSNYNE 1315
Db 409 AOLCGGNVYKFVTLHRPKOGADTYLELGFKN-GTLAPGASTGNQLRLHNDWNSYAO 466
QY 1316 SDDYSYKGTQSFDDHFKALYHNGVLWNGTE 1347
Db 467 SGDYSFFKSNT-FKTTKRTILYDQGLKINGTE 497

RESULT 13
GUN2_BACSU STANDARD; PRT; 499 AA.
AC GUN2_BACSU
AD P10475;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
GN BGC OR GLD OR EGLS.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAP115;
RX MEDLINE=87066783; PubMed=3024130;
RA Mackay R.M., Lo A., Willick G., Zuker M., Baird S., Dove M.,
RA Moraneli F., Selligy V.;
RT "Structure of a Bacillus subtilis endo-beta-1,4-glucanase gene.";
RL Nucleic Acids Res. 14:9159-9170(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CK-2;
RX MEDLINE=95225656; PubMed=7710280;
RA Lindahl V., Aa K., Tronsmo A.;
RT "Nucleotide sequence of an endo-beta-1,4-glucanase gene from Bacillus
subtilis CK-2.";
RL Antonie Van Leeuwenhoek 66:327-332(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Rose M., Entian K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]

```

```

RP SEQUENCE OF 30-45.
RC STRAIN=CK-2;
RX MEDLINE=95225655; PubMed=7710279;
RA Aa K., Flensburg R., Lindahl V., Tronsmo A.;
RT "Characterization of production and enzyme properties of an
endo-beta-1,4-glucanase from Bacillus subtilis CK-2 isolated from
compost soil.";
RL Antonie Van Leeuwenhoek 66:319-326(1994).
CC -/- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
linkages in cellulose.
CC -/- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
HYDROLASES).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; Z29076; CAA82317.1; -
DR EMBL; X04689; CAA28392.1; -
DR EMBL; X67044; CAA47429.1; -
DR EMBL; Z73234; CAA97610.1; ALT_INIT.
DR EMBL; Z99113; CABI3696.1; ALT_INIT.
DR PIR; A26114; A26114.
DR HSSP; O85465; IA3H.
DR Subtilist; BG10437; bg1C.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; Glyco_hydro_F5.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal;
KW Complete proteome.
FT SIGNAL 1 29
FT CHAIN 30 499 ENDOGLUCANASE.
FT ACT_SITE 169 169 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 350 499 CELLULOSE-BINDING (BY SIMILARITY).
FT CONFLICT 283 283 S -> N (IN REF. 2).
SQ SEQUENCE 499 AA; 55287 MW; 8F735FF711B3EA2 CRC64;

Query Match 4.8%; Score 347.5; DB 1; Length 499;
Best Local Similarity 33.1%; Pred. No. 6,7e-12;
Matches 84; Conservative 42; Mismatches 107; Indels 21; Gaps 8;

QY 1099 TEQPSVPGSFTLNAAGETAKV----KLSWTASANAASVEYKRVSTVNGAFATVADSYGVS 1154
Db 260 TSDASGNGGVFLDQSKREWLYLDSKTISW-VWNWLSKQESSALKPGASKT-----G 311
QY 1155 SYTDVAVTADTYTYQ-VVAKNDACQTVSNATASAPKADTQPTTGLVLQYRTADTNVD 1213
Db 312 GWRLSLDSASGTFVRENILGTNDKTDIPET----PSKDKPTQENGISVQYRAGDSMNS 367
QY 1214 NNLNPHFQILNKGITISVPINELKIRYYITIDGREQTFNCYAVLSCSKLNGKLVKMDKA 1273
Db 368 NOIRPOLQIKNGNTVVDLKDVTARYWNAK-NKQNVDCDYAQICGCGNTHKFFVTLHKP 426
QY 1274 ATGADYILEVSNFSDAGVLASGGTGGIQTIRHKADWSNYNESDDYSYKGTQTSFDDHTK 1333
Db 427 KQGAADTYLELGFKN-GTLAPGASTGNQLRLHNDWNSYAOQSGDYSFFKSNT-FKTTKK 483
QY 1334 ATLYHNGVLWNGTE 1347
Db 484 ITLYDQGLKINGTE 497

RESULT 14
CIPA_CLOTHM

```

ID AC CIPA_CLOTH STANDARD; PRT; 1853 AA.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cellulosomal scaffolding protein A precursor (Cellulosomal
 GN glycoprotein S1/SL) (Cellulose integrating protein A) (Cohesin).
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40 AND 42-43.
 RC STRAIN=ATCC 27405;
 RX MEDLINE=93302508; PubMed=8316083;
 RA Gergröss U.T., Romaniec M.P.M., Kobayashi T., Huskisson N.S.,
 RA Domain A.L.;
 RA "Sequencing of a Clostridium thermocellum gene (cipA) encoding the
 RT cellulosomal SL protein reveals an unusual degree of internal
 RT homology.";
 RL Mol. Microbiol. 8:325-334(1993).
 RN [2]
 RP SEQUENCE OF 1820-1853 FROM N.A.
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface.";
 RL J. Bacteriol. 175:1891-1899(1993).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 184-321.
 RX MEDLINE=97238934; PubMed=9083107;
 RA Shimon L.J.W., Bayer E.A., Morag E., Lamed R., Yaron S., Shoham Y.,
 RA Frolow F.;
 RT "A cohesin domain from Clostridium thermocellum: the crystal
 RT structure provides new insights into cellulosome assembly.";
 RL Structure 5:381-390(1997).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 368-522.
 RX MEDLINE=97076134; PubMed=8918451;
 RA Torino J., Lamed R., Chirino A.J., Morag E., Bayer E.A., Shoham Y.,
 RA Steltz T.A.;
 RT "Crystal structure of a bacterial family-III cellulose-binding
 RT domain: a general mechanism for attachment to cellulose.";
 RL EMBO J. 15:5739-5751(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1213-1361.
 RX MEDLINE=98022914; PubMed=9402065;
 RA Tavares G.A., Beguin P., Alzari P.M.;
 RT "The crystal structure of a type I cohesin domain at 1.7-A
 RT resolution.";
 RL J. Mol. Biol. 273:701-713(1997).
 CC -!- FUNCTION: ACTS AS A SCAFFOLDING PROTEIN IN THE CELLULOSE. IT
 CC PROMOTES BINDING OF CELLULOSE TO THE CATALYTIC DOMAINS OF THE
 CC CELLULOLYTIC ENZYMES.
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE.
 CC -!- DOMAIN: CONTAINS 9 COPIES OF A DOMAIN (COHESIN) OF ABOUT 150
 CC RESIDUES THAT BINDS TO THE DOCKERIN DOMAIN BORN BY THE CATALYTIC
 CC COMPONENTS OF THE CELLULOSE.
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
 CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L08665; -; NOT_ANNOTATED_CDS.

DR EMBL; X67506; CAA47840.1; -;
 DR PIR; S36859; S36859.
 DR PDB; 1ANU; 23-JUL-97.
 DR PDB; 1AOH; 08-JUL-98.
 DR PDB; 1NBC; 26-SEP-97.
 DR InterPro; IPR001956; CBD_3.
 DR InterPro; IPR002102; Cohesin.
 DR InterPro; IPR002105; Dockerin_1.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00942; CBD_3; 1.
 DR Pfam; PF00963; Cohesin; 9.
 DR Pfam; PF00404; Dockerin_1; 2.
 DR ProDom; PD001947; CBD_3; 1.
 DR ProSite; PS00018; EF_HAND; UNKNOWN_1.
 DR ProSite; PS00448; CLOS_CELLULOSE_RPT; 2.
 KW Cellulose degradation; Cell wall; Glycoprotein; Repeat; Signal;
 KW 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 29 1853 CELLULOSE SCAFFOLDING PROTEIN A.
 FT DOMAIN 364 522 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 54 1694 9 X 150 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 29 182 1-1.
 FT REPEAT 183 322 1-2.
 FT REPEAT 560 704 1-3.
 FT REPEAT 724 866 1-4.
 FT REPEAT 889 1031 1-5.
 FT REPEAT 1054 1196 1-6.
 FT REPEAT 1219 1361 1-7.
 FT REPEAT 1384 1526 1-8.
 FT REPEAT 1548 1690 1-9.
 FT DOMAIN 1791 1846 2 X 24 AA APPROXIMATE REPEATS.
 FT REPEAT 1791 1814 2-1.
 FT REPEAT 1824 1847 2-2.
 FT CONFLICT 1615 1615 A -> AA (IN REF. 1).
 SQ SEQUENCE 1853 AA; 196831 MW; 3ABDDC03ABFC5372 CRC64;
 Query Match 4.7%; Score 339.5; DB 1; Length 1853;
 Best Local Similarity 31.6%; Pred. NO. 1.2e-10;
 Matches 8; Conservative 48; Mismatches 95; Indels 45; Gaps 9;
 QY 1101 QPSVPGSFTLNAAGETAKVKLSWTASANAASY-----EVRKSTVENGAF 1145
 Db 264 EDSTGAYATKDGVPKIRA--TVKSSAFGYITFDEVGGFADNDLVEQKVSFDGGVNV 321
 QY 1146 TVASDVYGSSTDTAV---TADTTYTYQVYAKNDAGTQVNTASAPKADTQPTTG-LV 1201
 Db 322 GNATPTKGATPTATPTKATATPTPSVPTNTPTNTPTANT-----PVSONLK 370
 QY 1202 LOVRTADTNVNDNHLNPHFOILANKGTISVPINELKIRYYTIDGDEOTFCNDYAVL--- 1258
 Db 371 VEFYNSPSTTNSINPQFVNTGSSAIDLSKLTIRYYTYVDGKDDQTPWCDAHAIGS 430
 QY 1259 -----SCSKLNGKLVKMDKAATGADYILEVSNFNSDAGVLASGGSTGGIOTRIHKADWSN 1312
 Db 431 NGSYNGITSNVKGTFFVKMSSSTNNADTYLEISFT--GGTLEPGAHV-QIOGREAKNDWSN 487
 QY 1313 YNESDDYSYKGTQTSFDDHTKATLYHNGVLWGTE 1347
 Db 488 YTSQNDYSFK-SASQFVENDQVQVAYLNGVLWGKE 521
 RESULT 15
 CIPB_CLOTH STANDARD; PRT; 772 AA.
 ID CIPB_CLOTH
 AC Q01866;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cellulosomal scaffolding protein B (Cellulosomal glycoprotein S1/SL)
 DE Cellulose integrating protein B (Fragment).
 GN CIPB.
 OS Clostridium thermocellum.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 13, 2002, 00:58:37 ; Search time 91.54 Seconds
(without alignments)
2551.269 Million cell updates/sec

Title: US-09-784-554B-4

Perfect score: 7165

Sequence: 1 MKAHSSSIWSKRSKWLPPV.....HTKATLYHNGVLVWGTEPAN 1350

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1705	23.8	1601	2 P71140	P71140 clostridium
2	1534	21.4	606	16 Q977V3	Q977V3 clostridium
3	1524	21.3	1751	2 Q9A0G4	Q9A0G4 caldicellul
4	1106	15.4	632	2 Q22743	Q22743 ruminococcus
5	1082.5	15.1	808	2 Q334F9	Q334F9 ruminococcus
6	1037	14.5	398	2 Q30654	Q30654 dictyoglomu
7	920.5	12.8	911	2 P77847	P77847 caldicellul
8	801.5	11.2	591	2 Q9F1T9	Q9F1T9 clostridium
9	711.5	9.9	589	2 Q3REK7	Q3REK7 clostridium
10	549	7.7	921	2 Q9L8L8	Q9L8L8 caldicellul
11	540.5	7.5	930	2 Q9RFX5	Q9RFX5 caldicellul
12	525.5	7.3	170	2 Q9RFX6	Q9RFX6 caldicellul
13	521.5	7.3	209	2 Q9A0G9	Q9A0G9 caldicellul
14	477	6.7	997	2 Q9Z4L1	Q9Z4L1 bacillus sp
15	439.5	6.1	1770	2 Q9X3P5	Q9X3P5 caldicellul
16	439	6.1	1779	2 O52374	O52374 caldicellul

17 434.5 6.1 1915 2 Q9RPL0
18 423.5 5.9 1426 2 Q9X3P6
19 401 5.6 924 2 O06842
20 400.5 5.6 996 2 Q9A0H0
21 395.5 5.5 1000 2 Q24820
22 391.5 5.5 1711 2 P96311
23 377.5 5.3 502 2 Q9ADL2
24 361.5 5.0 261 2 Q9AQG7
25 360.5 5.0 386 2 O53324
26 352.5 4.9 499 2 Q45532
27 351 4.9 887 2 Q9L3J8
28 348.5 4.9 499 2 Q93TJ6
29 348.5 4.9 508 2 Q93LD0
30 346.5 4.8 501 2 O83012
31 338.5 4.7 499 2 O52731
32 312 4.4 2316 2 Q9FDJ9
33 301.5 4.2 473 2 Q9RK75
34 291 4.1 1483 16 Q977Y4
35 289.5 4.0 486 2 Q45430
36 289 4.0 362 2 O83011
37 286 4.0 92 2 Q9L8L9
38 283 3.9 362 16 O05512
39 270.5 3.8 1546 2 Q45996
40 265 3.7 367 2 Q9AG23
41 260.5 3.6 1230 2 O59325
42 254.5 3.6 1162 2 O82830
43 242 3.4 660 2 Q9L3J2
44 238.5 3.3 321 2 Q46392
45 225 3.1 1010 2 Q9XCV5

ALIGNMENTS

RESULT 1

P71140
ID P71140 PRELIMINARY; PRT; 1601 AA.
AC P71140;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENDOGLUCANASE.
GN CELJ.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_FaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FI;
RX MEDLINE=96422002; PubMed=8824619;
RA Ahsan M.M., Kimura T., Karita S., Sakka K., Ohmiya K.;
RT "Cloning, DNA sequencing, and expression of the gene encoding
RT clostridium thermocellum cellulase cellJ, the largest catalytic
RT component of the cellulosome."
RL J. Bacteriol. 178:5732-5740(1996).
DR EMBL: D83704; BAA12070.1;
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR004197; celd.N.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR001701; Glyco_hydro_9.
DR InterPro: IPR000601; PKD_domain.
DR Pfam: PF02927; celd.N.1.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF00759; Glyco_hydro_9; 3.
DR Pfam: PF00801; PKD; 1.
DR SMART; SM00089; PKD; 1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN1.
DR PROSITE; PS00448; CLOS_CELLULOsome_RPT; 1.
DR PROSITE; PS50093; PKD; 1.
SQ SEQUENCE 1601 AA; 178058 MW; 31E85D77F8642565 CRC64;

```

Query Match      23.8%; Score 1705; DB 2; Length 1601;
Best Local Similarity 45.5%; Pred. No. 2.7e-78;
Matches 372; Conservative 82; Mismatches 237; Indels 126; Gaps 18;

Qy 29 GALTPTVVGQTAQTVTKVDSKDRKPSPIYGTNQLAGDENITARRLGGNRTGY 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 766 GVTPNPT---EPKVVVDIRDTSAERKPSPIYGTNQLAGDENITARRLGGNRTGY 819
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 89 WNNMNSAGSDWQSSDLSNAGLTKAECEKPGAVATSEHDOSI-KOGTYSLATLPM 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 820 WNNMNSAGSDWQSSDLSNAGLTKAECEKPGAVATSEHDOSI-KOGTYSLATLPM 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 148 AGYVAKDNGSVQSEKAPSAWNEVYVNAKNAFQQLDQKQVYADEFVNFVKYGV 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 880 AGYVADNGSVQSEKAPSAWNEVYVNAKNAFQQLDQKQVYADEFVNFVKYGV 207
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 208 ASTGTGKGYSLDNEPALWSTHPRHGEKVGAKELVDRSVLSKAKAVDAAGAEITGPV 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 940 ASTGTGKGYSLDNEPALWSTHPRHGEKVGAKELVDRSVLSKAKAVDAAGAEITGPV 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 268 LYGFCAVKDQTAQTVTKVDSKDRKPSPIYGTNQLAGDENITARRLGGNRTGY 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1000 LYGFCAVKDQTAQTVTKVDSKDRKPSPIYGTNQLAGDENITARRLGGNRTGY 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 328 GIRIT--NEVGNDETAKKARQAPRTLDPTYKEDSWIAQWNSFELPLRLKQSVDKYYP 385
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1060 GERICFADPRNIETNARLQAPRTLDPTYKEDSWIAQWNSFELPLRLKQSVDKYYP 385
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 386 GTKALTEYSYGGENDISGGIMADVLGILGKNDVYVWYKWKLDGANNVYSAAYKLRYN 445
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1120 GTKALTEYSYGGENDISGGIMADVLGILGKNDVYVWYKWKLDGANNVYSAAYKLRYN 445
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 446 YDGSSTFGDISVHAQTSDIVNSVSHASVTDASYKELHLVVMKMSDMSAFDAQFDLSGET 505
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1178 YDGSSTFGDISVHAQTSDIVNSVSHASVTDASYKELHLVVMKMSDMSAFDAQFDLSGET 505
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 506 TYGSKVWFGDKNSQIKEAPITQISGNRTYTVPLTAYHIVLTA-----GNDT 556
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1238 NVTIGNWAFDRGSSNITQRTPIVNIKDNFTYTVPAALTACHIVLEAAEPVWYGLNDS 1297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 557 PVENPESFALR-----AEAGDKSTYLDASSGVVGS-----VQRAT 593
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1298 KVNADVIMLRYILGIIDNLTAAO---IYFD---GVNNSDYNIMKRYLLKAEIDIP 1351
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 594 Y--ENGPEAAVNASLVTSTYDNTV-----NGTSYVV-----KIT 1367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1352 YVPEQAQKAITFTSPEDVTDENVVFNASNSIDEDGTIAYWAFDGDGYEGTSTTPTIT 1411
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 628 AKTKT-----GTSASNVLKAVPRAPVDPDRYEADGTLGTVVSSGTF 673
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1412 YKYNPGTYKVLIVTDNQGASSFTATIKVTSATGNSKFNEDGTLGG--FTTSGT-- 1467
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 674 SGTGVTVTFNH-----AGDSLTMTIQAPTAGLY 701
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1468 NATGVVWTTKAPKRGRLKWTVSEGETAEKLDGGTIVVPGTWTTRIPSCAPI 1527
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 702 NLITGYSRPHDKRTNFSNLGKASGELYLWKATDFKE 738
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1528 AAIQPIYMPHTPDWSEVLWNSWTKGYTMV-KTDDWNE 1563
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
Q977Y3 ID Q977Y3 PRELIMINARY; PRT; 606 AA.
AC Q977Y3;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ENDOGLUCANASE A (ENDO-1,4-BETA-GLUCANASE) (CELLULASE A),
GN CAC0915.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

Query Match      21.4%; Score 1534; DB 16; Length 606;
Best Local Similarity 51.3%; Pred. No. 3.4e-70;
Matches 291; Conservative 85; Mismatches 175; Indels 16; Gaps 7;

Qy 12 KRKWLPPVVMACITIVGGALPTPT---VHGTAKTIVTIKVDTSKDRKPSPIYGTNQE 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 KRKWLPPVVMACITIVGGALPTPT---VHGTAKTIVTIKVDTSKDRKPSPIYGTNQE 68
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 69 LAGDENLTAARRLGGNRTYVWYKWKLDGANNVYSAAYKLRYN 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 FS-NAKVTARRLGGNRTYVWYKWKLDGANNVYSAAYKLRYN 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 129 SHDQSLKQGT-YSLVTLPMACYVAKDNGSVQSEKAPSAWNEVYVNAKNAFQQLD 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 AFHDSLWAGVSLVTLPMACYVAKDNGSVQSEKAPSAWNEVYVNAKNAFQQLD 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 188 KQNVYADEFVNFVKYKGYVASTGTGKGYSLDNEPALWSTHPRHGEKVGAKELVDRS 247
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 TDGVSVMDEFVNFVKYKGYVASTGTGKGYSLDNEPALWSTHPRHGEKVGAKELVDRS 237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 248 VLSLKAQKAVDAGAEITGPVLYGFGAYKDLQTAQTVTKVDSKDRKPSPIYGTNQE 307
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 TQLAQVVKKIDPAEATFGPALFSAFNSPQSSVSKGYNQWFIYDLDMKNSDA 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 308 EGRRLDVFVHWYPMAGGGIRI-TNEVGNDETAKKARQAPRTLDPTYKEDSWIAQW 366
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 AKRLLDALDLHWYPMAGGGIRI-TNEVGNDETAKKARQAPRTLDPTYKEDSWIAQW 357
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 367 SEPLPLRLKQSVKDYKPYGTGTLALTEYSYGGENDISGGIMADVLGILGKNDVY 426
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 KWGLPLIPKVSIDKPYGTGTLALTEYSYGGENDISGGIMADVLGILGKNDVY 417
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 427 KLKDGANNVYSAAYKLRYNVDGKSTFGDISVHAQTSDIVNSVSHASVTDASYKEL 486
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 ECNSDKNNYVQSAFNLNNYDGNNSKYGTDVCKDTSIDNNSTYASVTSNNGNMKDI 477
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 487 MNKSMDSAFADQLDSETTYGSGKVGWFGDKNSQIKEAPITQISGNRTYTVPLTAY 546
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 MNKNYTDNISFNFNVSNNKNTSGQVWFGDSSNNTKRDDVSSISGNKFTYKIPALTAV 537
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 547 HIVLTAGNTPVENPESFALRAEADG 573
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 538 HIVLTAGNTPVENPESFALRAEADG 557
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
Q9AQQ4 ID Q9AQQ4 PRELIMINARY; PRT; 1751 AA.
AC Q9AQQ4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

```

```

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CELE.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Thermoaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOK7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D.; Reeves R.A.; Farrington G.K.; Anderson P.; Williams D.P.;
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078042; AAK06394.1; -
DR HSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR005566; Lipocln_cytfabp.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam: PF00942; CBD_3; 1.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR ProDom; PD001947; CBD_3; 3.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 1751 AA; 192176 MW; 60178CBF3C00BE95 CRC64;

Query Match 21.3%; Score 1524; DB 2; Length 1751;
Best Local Similarity 53.5%; Pred. No. 4.9e-69;
Matches 289; Conservative 76; Mismatches 151; Indels 24; Gaps 6;

QY 32 PPTVTVHGOQT-----AKVTIKVDTSKDRKIPSPYIYGTNOELAGDENLITARRLGGNRT 86
DB 1215 PPTVTVTPTPAVTPDVKISIDTSRGRTRKISPIYIGANQIQGVVH-PARRLGGNRLT,1273

QY 87 GYNWENNNSAGSDWQSDSYLDCNAGLTAKCEKPCAVATSFHDQSKOGTYSLVTL 146
DB 1274 GYNWENNNSAGSDWQSDSYLDCNAGLTAKCEKPCAVATSFHDQSKOGTYSLVTL,1333

QY 147 MAGYVAKDNGSVQSEKAPARNEVNVNNAKNAPELQDPLKDNQVADFEVNLVKYG 206
DB 1334 MGYVAKDNGTYSVSETPAPRAVEKFKDGLSLOPDVNDNFVYVDFEINTLINKYG 1393

QY 207 VASTKTGKGYSLDNEPALWSHTPHRIGEKVGAKELVDRSVLSKAAYDAGAEIFGP 266
DB 1394 RSSSATGKGYSLDNEPALWSHTPHRIGEKVGAKELVDRSVLSKAAYDAGAEIFGP 1453

QY 267 VLYGFCAYKDLQADPNSVKNYSWFVDYLDQWRLSSQAEGRLLDVFVHWYPEAMG 326
DB 1454 ASYGFGYLTLDQADPNSVKNYSWFVDYLDQWRLSSQAEGRLLDVFVHWYPEAOV 1513

QY 327 GGIRITNEVGNDETCK-----ARMAQPTLWDPYTK-----EDSWIAOWNSEFLPL 372
DB 1514 GGVRICFD-GENSTRDVALARMQAPRTLWDPYTKTQKGITAGENSINQWPEFLPL 1572

QY 373 LPRLKQSDVKYYPGKLTALTEYSYGGENDISGGIADVLGILGKNDVIMANYWKLK 432
DB 1573 LPRNKADIKYYPGKLTALTEYSYGGENDISGGIADVLGILGKNDVIMANYWKLK 1629

QY 433 NNYVSAAYKLYRNDGKSTGDSVHAQTSVSNVSHASVTDASYKELHLVNMKSMO.492
DB 1630 GSYAQAAAYLYRNDGKSTGDSVHAQTSVSNVSHASVTDASYKELHLVNMKSMO.492

QY 493 SAFDAQFOLSGEYTGSGKVGFDKNSOIKEAAPITQISGNRRTYTPVPLTAYHIVLTA 552
DB 1690 RLKAEIKNNWTRVYTGGEIGFDSSTSOIKRMGVLSIQNNNTTIEVPNLTVTHVLT 1749

QY 493 SAFDAQFOLSGEYTGSGKVGFDKNSOIKEAAPITQISGNRRTYTPVPLTAYHIVLTA 551
DB 470 EAENAVIDLNASDKYKSAAYVAVYGDNDQVRLDIVDKVKNVVELPAFSAAMVYVS 529
QY 552 AGNDTPVENPESPALRAEADGKSYLDSAGGVGVSVORATYENGPPFAAVASNLVETSY 611
DB 530 --DD-----AAAFDGEKIYEK----- 544
QY 612 TDNTVNTNGTSYYKIITAKTKGTGSASNVL--KAVPRAPVDGPDPRYAEADGTLKGTVV 669

```

RESULT 4

```

Q52743
ID Q52743 PRELIMINARY; PRT; 632 AA.
AC Q52743;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CELB.
GN CELB.
OS Ruminococcus flavefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Ruminococcus.
OX NCBI_TaxID=1265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FD-1;
RX MEDLINE=96025594; PubMed=8590402;
RA Vercoe P.E.; Finks J.L.; White B.A.;
RT "DNA sequence and transcriptional characterization of a beta-glucanase
RT gene (celB) from Ruminococcus flavefaciens FD-1.";
RL Can. J. Microbiol. 41:869-876(1995).
DR EMBL: U08621; AAA95959.1; -
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR001000; Glyco_hydro_10.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.
SQ SEQUENCE 632 AA; 69491 MW; CA02A5F93294769C CRC64;

Query Match 15.4%; Score 1106; DB 2; Length 632;
Best Local Similarity 36.6%; Pred. No. 2.1e-48;
Matches 247; Conservative 109; Mismatches 218; Indels 100; Gaps 16;

QY 23 CTIIVG-GALPTPTVHGOQTAKVTIKVDTSKDRKIPSPYIYGTNOELAGDENL--TARR 79
DB 16 CTLSAPLGLVP-ETKYNAAAGGFDMMIKVDLKGKKEISPLIYGVNYYTDLKSVKTTAVR 74

QY 80 LGGNRTMGYNWENNNSAGSDWQSDSYLDCNAGLTAKCEKPCAVATSFHDQSKOGT 139
DB 75 QGNNRTMGYNWENNNSAGSDWQSDSYLDCNAGLTAKCEKPCAVATSFHDQSKOGT 126

QY 140 -YSLVTLPMAGYVAKDNGSVQSEKAPARNEVNVNNAKNAPELQDPLKDNQVADFEV 198
DB 127 DYKWTTLQMAGYVADKDGTVKEDVAPSKRWNEVFTKGAPEADPLTDGVTYVNDYV 186

QY 199 NFLVKYGVASTGTGKGYSLDNEPALWSHTPHRIGEKVGAKELVDRSVLSKAAYD 258
DB 187 NYIINKLQDSQSTGTGQYSLDNEPALWSHTPHRIGEKVGAKELVDRSVLSKAAYD 246

QY 259 AGAEIFGPLYGFCAYKDLQ-----TAPDMSVKNYSWFVDYLDQWRLSSQAEGRKL 312
DB 247 PKAEIFGPLYGFCAYKDLQ-----TAPDMSVKNYSWFVDYLDQWRLSSQAEGRKL 305

QY 313 LDVDFVHWYPEAMGGGIRITNEVGNDETCKARMAQPTLWDPYTKEDSWIAOWNSEFLPL 372
DB 306 LDVLDIHYSESARTG-----AEDRVQSVRTLYEEGFSSENSWIGQWCMQNPVI 353

QY 373 LPRLKQSDVKYYPGKLTALTEYSYGGENDISGGIADVLGILGKNDVIMANYWKLK 432
DB 354 LPTIKKSIDYYPGKLTALTEYSYGGENDISGGIADVLGILGKNDVIMANYWKLK 409

QY 433 NNYVSAAYKLYRNDGKSTGDSVHAQTSVSNVSHASVTDASYKELHLVNMKSMO 492
DB 410 EPIIISINLYTYDGGKCGFDTLIPASTEDYSKSTTAAVNDGDESKVTYVITNKNMT 469

QY 493 SAFDAQFOLSGEYTGSGKVGFDKNSOIKEAAPITQISGNRRTYTPVPLTAYHIVLTA 551
DB 470 EAENAVIDLNASDKYKSAAYVAVYGDNDQVRLDIVDKVKNVVELPAFSAAMVYVS 529
QY 552 AGNDTPVENPESPALRAEADGKSYLDSAGGVGVSVORATYENGPPFAAVASNLVETSY 611
DB 530 --DD-----AAAFDGEKIYEK----- 544
QY 612 TDNTVNTNGTSYYKIITAKTKGTGSASNVL--KAVPRAPVDGPDPRYAEADGTLKGTVV 669

```

```
Db 545 -----KVTEEFKDPSSMINKNGVVEIPITDPE--QVSKIVINGDVTSSA 589
Qy 670 GTGFGSGTGYVNFH 683
Db 590 GSGWA-TARLCMSH 602

RESULT 5
Q934F9 PRELIMINARY; PRT; 808 AA.
AC Q934F9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CELLULOSE PRECURSOR.
GN ENDB.
OS Ruminococcus flavefaciens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae.
OC Ruminococcus
OX NCBI_TaxID=1265;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17;
RX MEDLINE=21454991; PubMed=115711138;
RA Rincon M.T., McCrae S.I., Kirby J., Scott K.P., Flint H.J.;
RT "EndB, a newly identified family 44 cellulase from the rumen
cellulolytic bacterium Ruminococcus flavefaciens 17, binds to
cellulose via a novel cellulose binding domain and to a 130kDa
R. flavefaciens protein via a dockerin domain.";
RL Appl. Environ. Microbiol. 67:4426-4431(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=17;
RA Rincon M.T.;
RL Thesis (2001), Department of Agriculture, University of Aberdeen,
RL Aberdeen, United Kingdom.
DR EMBL; AJ298117; CAC83072.1; -.
KW Signal.
FT SIGNAL 1 32 POTENTIAL.
FT CHAIN 33 808 ENDB PROTEIN.
SQ SEQUENCE 808 AA; 87750 MW; 8E1CDF28948ECEA0 CRC64;

Query Match 15.1%; Score 1082.5; DB 2; Length 808;
Best Local Similarity 33.8%; Pred. No. 4.6e-47;
Matches 263; Conservative 120; Mismatches 263; Indels 133; Gaps 23;

Qy 19 VMACTII--VGGALPTPTVHGGTARTVPIKVDTSKDRKPISPYIYGTNOELAG--DEN 74
Db 12 VVSGCTWTAPVNG-LPRAT-VNAGGYDMNVTVDLKGEKKAISPLIYGVNQYITDRLDYK 69
Qy 75 LTARLGGNRTGYNWNENNSNAGSDMSYSDYLCNAGLTKAECEKPGAVATSPHDQS 134
Db 70 TTAVRQGNRMTAYNWTETNAGSDMKHSSDNLSDS-----DDPADCVQLSKQA 121
Qy 135 LKQG-TYSLVTLPMAGYVAKDNGSVQESKAPSAKRWENNAKNAPFQLOPLDKNQY 193
Db 122 AKYNNVKKLTQLAGVTSADKNGPVSEAEKAPSDRWNVKVLTKNAPFADTDLDTGVVY 181
Qy 194 ADEFVNLVKYGVASTKTKGVKYSLDNEPALMSHTHPRIHGBKVGAKELVDRSVLSKA 253
Db 182 MDEVVYIINKLGDSSQSAEGIQGYSLDNEPVLWNTDSRHPDPVPIELGSKSVEMAKA 241
Qy 254 AKAVDAGAEIFGVLGVFGAYKDL--QTAPDNNSVK--GNYSWFVDYLDQMRLSQAE 308
Db 242 VKKLDPAEYFGALYGYTAFDRLDDDDDAHEWEELKANNVHWYLDYLDHMHKASEEN 301
Qy 309 GKRLDVEDVHWYPEAMGGGIRITNEVGNDETKKARMQAPRTLDPTYKEDSKIAQWNE 368
Db 302 GARLLDVLDIHYSERKGI-----EDRLQSVRTLYEPFGFSENSWIGOWCME 349
Qy 369 FLPLPLRKQSDVKYYPGTKALTEYSYGGENDISGGIAMDVLGLTKLGNVYMANWK 428
```

```
Db 350 NVPIPTIOKSIDTYPGTKLGISEYFNFGGDDASGTIAQAALGCGYADQGYVFASLW-- 407
Qy 429 KDGANNVSAAYKLYRNYDGKSTFGDISVHAQTSDIVNSSVSHASVTDASYKELHLVWN 488
Db 408 --GGEFFILSGQLTYNYDGGKCGFGDTLIPASTGDVSKSSYAAVNAKDDSKVTVMVTN 465
Qy 489 KMSDAFDAQFDL-SGETTYGSGKVMGFDKNSQIKAAAPITQISGNRFYTVPPPLTAYH 547
Db 466 KDLKENENAVIDLNRNADKSKYKSAAYVAVFGDSEELRLIDI IKVDKDNKVKTELPAPSAAM 525
Qy 548 IVLTAGNDTPVENPESFALRAAGDGKSIYLDASSGVGVSYVORATYENGPPAAVASNLV 607
Db 526 VVWSD-----QADAFDLKTYEETKT-----ETKVE----- 552
Qy 608 ETSYTD-TNWTNGTSYIYKITTAKTTGTSASNVLKAVPRAPVDGDPDYEADGTLKGTVV 666
Db 553 ---FTDIESWTNDKGF---VVVPIEDAEHLKII-----INGAVT 586
Qy 667 ESSGTGFGSGTGYVNFHNGDSL/TMTIQAP-----TAGLYNLTYIGRSPHDKRTNFSL 720
Db 587 SSAGSSWA-----TAGCAVCMNVTAKGSGFWTYKSYNPLGSKS-----SATVKF 632
Qy 721 NG---KASGELVLWKTADPKET-SGSKVLLNAGANTIGFETGCGWYDIDYVKLEPAADR 775
Db 633 DGIFTKETGEGADKVKEDLEATVAGKVELQK-----WMDASEKSESEEDK 679

RESULT 6
O30654 PRELIMINARY; PRT; 398 AA.
AC O30654;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-MANNANASE (EC 3.2.1.78).
GN MANA.
OS Dictyoglomus thermophilum.
OC Bacteria; Dictyoglomus group; Dictyoglomus.
OX NCBI_TaxID=14;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RT46B.1;
RA Reeves R.A., Gibbs M.D., Bergquist P.L.;
RT "Sequence of a beta-mannanase from the extreme thermophile
Dictyoglomus thermophilum Rt46B.1.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF013989; AAB82454.1; -.
DR InterPro; IPR000805; Glyco_hydro_26.
DR Pfam; PF02156; Glyco_hydro_26; 1.
DR PRINTS; PR00739; GLHYDRLASE26.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 398 AA; 46492 MW; 0E564DEBA52CA18B CRC64;

Query Match 14.5%; Score 1037; DB 2; Length 398;
Best Local Similarity 50.6%; Pred. No. 3.5e-45;
Matches 202; Conservative 52; Mismatches 137; Indels 8; Gaps 5;

Qy 700 LYNLTIGRSPHDKRTNFSNLNGASGELVLWKTADPKETSGGKVLNAGANTIGFETGW 759
Db 1 MHELIIGYAAPGYKENSILYNGEFTQNVKFPQSQKFTTVYAGLIPLKNGKNTISIKSW 60
Qy 760 GWYDIDYVKLEPAADRPHPHAYTKTLINPNATVEAKALMNYLDQYGNMLSGQEDMPEID 819
Db 61 GWFLLDYFKIK-KAEIPTMNPNTKLVTPNPSKAQKLDYLVSIYGYKTLVLSQGYKDAF 119
Qy 820 WLOANVGKPKAIALDLIDYSPSRAEHCLSSSTETEKALEWDKQGGIVTFAHWNAPGLI 879
Db 120 WLNITDFFKAPCGFDMMDYSPSRVERGASSRDEDAIDMNMGGIVQFOHWNAPKGLY 179
Qy 880 DTQKEMWRGFGVADSTTDIEYAMNHPESDYKLLIRIDIVIAQGLKLODAKVPVLPFRP 939
```


Db 180 DTPGKWRGFTYNTATSFDEYALNHPSEYDKLIIRDIDAIIVOLKRLQEAQVPIILWRP 239
QY 940 LHEAGKWFVWAGKGPVKKLYLHMDRLTNVHKLNNLIWVNSV-APD---WYPCDEY 995
Db 240 LHEAGKWFVWAGKGPVKKLYLHMDRLTNVHKLNNLIWVNSV-APD---WYPCDEY 995
QY 996 VDLSFDSYPOAGDYSPQIAKYEDVLTKDKKLVCHERRKRTYPPGSGDEGVSSPLELYR 1055
Db 300 VDIVGADILYKDYSPSTGMEYINVLKFGKLVALTENGIIIPDP--DLMEQKAYVW 357
QY 1056 YVWGFLEFRTQTQSLEHLKLVYVHNHNVITLEKLPNLK 1094
Db 358 FMTWSGFENDPNKNEISHIKVFNHPFVITKDLP-NLK 395
RESULT 7
ID P77847 PRELIMINARY; PRT; 911 AA.
AC P77847;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-1,4-MANNANASE (EC 3.2.1.78).
GN MANA.
OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=44001;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RT8B.4;
RX MEDLINE=96291679; PubMed=8764509;
RA Gibbs M.D., Elinder A.U., Reeves R.A., Bergquist P.L.;
RT "Sequencing, cloning and expression of a beta-1,4-mannanase gene,
RT manA, from the extremely thermophilic anaerobic bacterium,
RT Caldicellulosiruptor Rt8B.4.";
RL FEMS Microbiol. Lett. 141:37-43(1996).
DR EMBL; U39812; AAC44232.1;
DR InterPro; IPR000805; Glyco_hydro_26.
DR Pfam; PF02156; Glyco_hydro_26; 1.
DR PRINTS; PR00739; GLYDRLASE26.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 911 AA; 103324 MW; AD4A7FF0C1C8DF0B CRC64;

Query Match 12.8%; Score 920.5; DB 2; Length 911;
Best Local Similarity 31.4%; Pred. No. 9.6e-39;
Matches 244; Conservative 97; Mismatches 261; Indels 175; Gaps 23;
QY 379 SVDKYYPGKTLALTEYSYGGENDISGGIAMADVLGILGKNDVY-----MANYWKLKGA 432
Db 237 AVEDEFTGVKLVYPEQ-----GASFVYNEIDTMGYKYSGDGF 275
QY 433 NNTV-SAAKLYKNYDGKSTGDISVHAQTSIDVNSVSHASVTDASYKELHVLVYNNKSM 491
Db 276 NKTKSLEYSDQLKRSNGNAL-----KVNASLAGTAFDEMIAIKLTDK 320
QY 492 DSADFADQFDLGGTYG-----SGKV-----WGFKNSSOI--KEAAPIT 529
Db 321 D---DKKDFSKYSTLEYLYIPNPKISGKLWASAINDPWQIIKDFTAIYNKDKNAIQ 377
QY 530 QISGNRFT-----YVPPLTAYHIVLTAGNDTPVENPESFALRAEAGDGKSIYLDAS 581
Db 378 KINGDQYAVIKSDNLYNNYTRANVLVRIAGSYKYTGCP-----IYIDNV 423
QY 582 SGVVGVSORATYENGPPFAASNLVETSYDTNVTNGTSYFYKITAKTGTTSASNLVK 641
Db 424 KLVAGKKV-----APKVKTTSSIPNKN-----YVRVKEAET 456
QY 642 AVPRAPVDGPDYAEADGTLKGVTVVSESSGTGCTGYTVNFNAGDSLMTWTQAPTAGLY 701
Db 457 -----AKDGWAYS--LEKENAKYSRGKYLILFGNMGNTLYDLIKPIKTYGY 500

QY 702 NLTIGYRSPHDDKRTNFSINLKASGELVLWKTADPKETSG-----GKVLIN 747
Db 501 IFTI-----SSSTLGMVNGSVDIWIWDELKGAAPVNVYKGFQEVVVMKKIYLT 550
QY 748 AGANTIGFETGWG-YDIDYVKLEP--AADPPHVAVTTLINPNATVEAKALMNYLVDOY 804
Db 551 OGEHTLSLQKSGYTIADYPTIEELVLANKNKISVDTKLVTNPNPNAQKLMKYLASIY 610
QY 805 GKNMLSGQE---DMPEIDWLQANVKKPAIALDLIDISPSRAEHGLSSTETEAIEWDK 861
Db 611 GEKILSGQSGDGRKEIQMIFDVTKRYPAVRGFDPMYSPSRVHGRTGTDVEEAIKWK 670
QY 862 OGGIVTFAWHNAPKGLIDTOGKEMWRCFYADSTTFDIEYAMNHPSESDYKLLIRIDVI 921
Db 671 SGGIVAFCHWNAPTGLIDQPKEMWRCFYTEATTFDICKAMDNPNSTFYKLLIRIDAI 730
QY 922 AGOLKKLQDAKVPVLFRLPHEAGKWFVWAGKPPVKKLYLHMDRLTNVHKLNNLIWV 981
Db 731 AEQLARLQKEGVPVLFRLPHEASGWFVWAGKGPPEYIKLKLMDRLVNLVHKLNNLIWV 790
QY 982 NNSVADPWYPGDEYVDILSFDSPQAGDYSPQIAKYEDVLTKDKKLVCHERRKRTYGP 1041
Db 791 WNGDAWYPGDNYVDIIGEDIYEKAQYSPYADRFKALAKYTNARKMIALTECGTIPDP 850
QY 1042 G-SDEGVSSPLELYRYVWV-GFLERRQTKOSLE-----HLKVVYNHNPVITLEKLP 1090
Db 851 AVLKQEGISWSW---FSWAGNFVMTGSKYNDENNDHMLRKIYNSDYVITKDLP 903
RESULT 8
QY 991T9 PRELIMINARY; PRT; 591 AA.
ID Q991T9;
AC Q991T9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MANNANASE 26B (EC 3.2.1.78).
GN MAN26B.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RA Kurokawa J., Hemjinda E., Arai T., Karita S., Kimura T., Sakka K.,
RA Ohmiya K.;
RT "Sequence of the Clostridium thermocellum Mannanase Gene man26B and
RT Characterization of the Translated Product.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB044406; BAB19050.1;
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF02156; Glyco_hydro_26; 1.
DR PRINTS; PR00739; GLYDRLASE26.
DR PROSITE; PS00018; EF_HAND; NOWNOWN_1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 591 AA; 67048 MW; 92AFCF808BD761EB CRC64;

Query Match 11.28%; Score 801.5; DB 2; Length 591;
Best Local Similarity 37.8%; Pred. No. 5.8e-33;
Matches 202; Conservative 83; Mismatches 186; Indels 63; Gaps 20;
QY 635 SASNLKAVPRAPVDGPDYAEADGTLKGVTVVSESSGTG-----FSGTGVTNF-HNAG 686
Db 23 SCINVSNAV-----LSDGDKYEFEDGIHKQAIYTDYQNEYGEVFDLTGSTCSFIAQKG 78
QY 687 DSLTWTQAPTAGLYNTIGYRSPHD-DKRTNF-SLNGASGELVLWKTADPKETSGKY 744
Db 79 TSTSVNVEVDKGLYFICVQYQYDKNKKVQVNLVNGVNGOGEISFPFPTLKWREISAGIV 138

```
Qy 745 LNLGANTIGFETGAGWYDIDYVLEPAADRPHP-AVTKTLINPNATVEAKALMNLVDQ 803
Db 139 KLNAGINNELESYWGTYFDYLVKPADESIVELKVPKLVNPNATKEAKALMNLVDI 198
Qy 804 YGKNMLSGQEDM-----PIDWLQANVKKPAIAALDLIDYSPSRAEHLGSLSTE- 852
Db 199 YGKHILSGQOEICGSHNYPGSEAEFTYIOEKTGLPAVRGDFMNY-----RGNGLMWDQ 254
Qy 853 -TEKAIWDKO-GGIVTFAMHWNAPKGLIDTQCKEMRGFYADSTFDIEYAMNHPESED 910
Db 255 CAERVIEWYKKGIPVCHWHFSPGDI---GKKADNSFYTESTFISRALTF-PCTEE 309
Qy 911 YKLLIRIDIVTAGLKKLQDAKVPVLPRLHEAGKFWGAKGPEVVKKLYILMHDRLT 970
Db 310 NIALNDITDITARKLKQVQDAGVPLVLPRLHEAGGFWGAKGPEPCVRLYLLYDKFT 369
Qy 971 NVHKLNLNIWVNS-----VAPDWYPGDEYVDILSFDSPQAGDYSPO-IAKYEDLVT 1022
Db 370 NEYGLNLIWVNTSYDSETSAAWYPGDVVDIIGYDKY-NAKDGKPNGSAISSTFYNLVK 428
Qy 1023 LGDKKKLVCHERKRTYPGSGDEGVSSPLELVRYMVW-GFLERRQTKOSLEHLKKVYNHP 1081
Db 429 LTNGKKLVAMTENDTIPRVSN--LVNEKAGWLXFCPWYGGWLTSEQNNPVDLWVEMQSD 486
Qy 1082 NVITLEKLPNTLNKYGITE-----QPSVPGSET---LNAAGETAKVKLS 1122
Db 487 YCITLDELPL-DLNKYPISDVEDSNPDPSPTQPKITYGDLNGDGKVNSTDILT 539

RESULT 9
Q9REC7 PRELIMINARY; PRT; 589 AA.
AC Q9REC7;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE MANNANASE A PRECURSOR (EC 3.2.1.78).
GN MANA.
OS Clostridium thermoceillum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=YS;
RX MEDLINE=20055634; PubMed=10589717;
RA Halstead J.R., Vercoe P., Gilbert H.J., Hazlewood G.P., Davidson K.;
RT "A family 26 mannanase produced by Clostridium thermocellum as a
RT component of the cellulosome contains a domain which is conserved in
RL mannanases from anaerobic fungi."
RL Microbiology 145:3101-3108(1999).
DE EMBL; AJ242666; CAB52403.1; -.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand_1.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF02156; Glyco_hydro_26; 2.
DR PRINTS; PR00739; GLHYDRASE26.
DR PROSITE; PS00448; CLOS_CELULOSOME_RPT; UNKNOWN_2.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL 1 30 POTENTIAL.
SQ SEQUENCE 589 AA; 66816 MW; 62F45CEB149978F9 CRC64;
```

Query Match 9.98; Score 711.5; DB 2; Length 589;
Best Local Similarity 35.08; Pred. No. 2.2e-28;
Matches 192; Conservative 79; Mismatches 183; Indels 95; Gaps 22;

```
Qy 635 SASNVKLVAPRAPDGPDRYEAEDGTLKGTWVSSGTG-----FSGTGVTVNF-HNAG 686
Db 23 SCINVSNAV----LSDGDKYEEFDGIHKGAIYTDYQNEYGEVFDLTGTCSEFIAQK 78
```

```
Qy 687 DSLTMTIQAPTAGLXNLITIGYRSPHD-DKRTNF-SLNGKASGELVLMKTADPKETSGKV 744
Db 79 TSTSVNVEVDKEGLYEIFCYVQPYDKKKVQYLVNNGVNOQEISFPFTLKKREISAGIV 138
Qy 745 LNLGANTIGFETGAGWYDIDYVLEPAADRPHP-AVTKTLINPNATVEAKALMNLVDQ 803
Db 139 KLNAGINNELESYWGTYFDYLVKPADESIVELKVPKLVNPNATKEAKALMNLVDI 198
Qy 804 YGKNMLSGQEDM-----PIDWLQANVKKPAIAALDLIDYSPSRAEHLGSLSTE- 852
Db 199 YGKHILSGQOEICGSHNYPGSEAEFTYIOEKTGLPAVRGDFMNY-----RGNGLMWDQ 254
Qy 853 -TEKAIWDKO-GGIVTFAMHWNAPKGLIDTQCKEMRGFYADSTFDIEYAMN----- 904
Db 255 CAERVIEWYKKGIPVCHWHFSPGDI---GKKADNSFYTESTFISRALTFCTRKI 310
Qy 905 -----HPSEDKYKLLIRIDIVTAGLKKLQDAKVPVLPRLHEAGKFWGAKGP 955
Db 311 LHCLTISTWPPEAQ-----AGSGCEF-----VLFRPLHEAGGFWGAKGP 352
Qy 956 EPVKKLYILMHDRLTNVHKLNLNIWVNS-----VAPDWYPGDEYVDILSFDSPQAGDYS 1011
Db 353 EPCVRLYRLLYDKYNTNEYGLNLIWVNTSYDSETSAAWYPGDVVDIIGYDKY-NAKDGK 411
Qy 1012 PQ-----IAKYEDLVTLGDKKKLVCHERKRTYPGSGDEGVSSPLELVRYMVW-GFLERRQ 1066
Db 412 PNGSAISSTFYNLVKLTNGKKLVAMTENDTIPRVSN--LVNEKAGWLXFCPWYGGWLTSE 469
Qy 1067 TKOSLEHLKKVYNHPNVTLEKLPNTLNKYGITE-----QPSVPGSET---LNA 1113
Db 470 QNNPVDLWVEMQSDYCIITDELPL-DLNKYPISDVEDSNPDPSPTQPKITYGDLNGD 528
Qy 1114 GETAKVKLS 1122
Db 529 GKNVSTDILT 537

RESULT 10
Q9L8L8 PRELIMINARY; PRT; 921 AA.
AC Q9L8L8;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE BETA-1,4-XYLANASE XYNA PRECURSOR.
GN XYNA.
OS Caldibacillus cellulosovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldibacillus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20519260; PubMed=11065373;
RA Sunna A., Gibbs M.D., Bergquist P.L.;
RT "A novel thermostable multidomain 1,4-beta-xylanase from
RT Caldibacillus cellulosovorans' and effect of its xylan-binding domain
RT on enzyme activity."
RL Microbiology 146:2947-2955(2000).
DR EMBL; AF200304; AAF61649.1; -.
DR HSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBD_6.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR002965; P_rich_extensin.
DR Pfam; PF02018; CBD_6; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDRASE10.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR ProDom; PD001947; CBD_3; 2.
KW Signal; Xylan degradation; Hydrolase; Glycosidase.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 921 BETA-1,4-XYLANASE XYNA.
SQ SEQUENCE 921 AA; 102380 MW; C5DDDLIA7F7567413 CRC64;
```

[illegible]

QY 705 IGRSPHDDKRTNLSN-----GKASGELVLMKTKETSGGKVL-LNAGANTIGFE 756
Db 485 VGINPHAWYDRRLSSSQITSGCANAVRVLNCGRWTKIPASEVADIISQARTLGYR 544
QY 757 -----TGWGVYDIDYVLEPAADRPHTKTLINPNATVEAKALMNYLDQYCKN 807
Db 545 AVLEVHTDTG-----EDAA-----ACSM-----TAVNYWIEL--KN 577
QY 808 MLSGDEMPEDLQANVKKPATAALDLIDYSPSRAEHLGSLSTETEKALWDKQGGIVT 867
Db 578 VLAGOE-----NFVIVNIGNEP-----YGNNNYQNWY--TDRNAVQALRNAGI-- 619
QY 868 FAWHNAKGLIDTQGEKWRGFEYADSTTFDIEYAMNHPESDYKLLRIDIDVIAGLKK 927
Db 620 -----NTIMWDAPNA-----GQDSFTWRD----- 640
QY 928 LODAKVPVFLRPLHEAEGKFWGAKGPEPVKKYILMHDRLTNVHKLNLWNNWVAP 987
Db 641 ----NAPTIF-----NADPQNLVFSIH----- 659
QY 988 DWYPCDEYVDILSPDSYPOAGDYSQYAKYEDLVTLGDKKKLVCHERKRYTPGDSDEGV 1047
Db 660 -----MYGVYDITAEVOSYIESF-----VNRGLPLVIGEF-----GHMHSDDG 697
QY 1048 SSPLLEVRYMVGFLERROTQKSLHLKKVYNHPNVITLEKLPNLTNYGITEQPSVPS 1107
Db 698 PNEQAIQV-----AKQYNI-----GL 714
QY 1108 FTLNAGETAKVKLSWTASANAAYEYKRVSTVENGAFATVADSYGSSYTDVATADTTY 1167
Db 715 F-----GWSWNGSGVGEYLDVMTNFNANSPTA---WGTFWRTNAGTSTS- 757
QY 1168 YQVQVAKNDAGQTVSNASAP-KADTQOPTTG--LVLOYRTADTNVNDHNLPHFQILN 1224
Db 758 -----PTPTPTPTPTPTPTPTPSAGGNLVVQYRAADTNATDNLKPHFRIVN 806
QY 1225 KGTISVPINELKIRYYTIDGROTFNCYAVLSCKLNGKLVKMDKAAATGADYLYEVS 1284
Db 807 RGTSSVPLSELTYRWYTVGDKQVFNCDWAQVGCNSLNGSFVKLSTGRTGADYIEIT 866
QY 1285 FNSDAGVLASGGSTGGIOTRIHKADWSNYESDDYSYGTOTGFDHHTKATLYHNGVLVW 1344
Db 867 FTSAGSLAPGASSGDIQVRINKNDWTNYEANDYSYDPTKTSFADNWRVTLYRNGQLVW 926
QY 1345 GTE 1347
Db 927 GVE 929

RESULT 12
Q9RFX6 PRELIMINARY; PRT; 170 AA.
AC Q9RFX6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEICAL 18.5 KDA PROTEIN (FRAGMENT).
OS Caldicoccus cellulosivorans
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Alicyclobacillus group; Caldicoccus.
OX NCBI_TaxID=74586;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20120520; PubMed=10653733;
RA Sunna A., Gibbs M.D., Chin C.W.J., Nelson P.J., Bergquist P.L.;
RT "A gene encoding a novel multidomain beta-1,4-mannanase from
Caldicoccus cellulosivorans and action of the recombinant enzyme on
kraft pulp."
RL Appl. Environ. Microbiol. 66:664-670(2000).
DR EMBL; AF163837; AAF22273.1;
DR HSP; O06851; INBC
DR InterPro; IPR001956; CBD_3.
DR Pfam; PF00942; CBD_3; 1.

DR ProDom: PD001947; CBD_3; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 170 AA; 18493 MW; 7AC9D33F44E3A0B4 CRC64;

Query Match 7.3%; Score 525.5; DB 2; Length 170;
Best Local Similarity 57.1%; Pred. No. 1.1e-19;
Matches 97; Conservative 27; Mismatches 41; Indels 5; Gaps 2;

QY 1180 TVSNTASAPKADTQOPTTG--LVLOYRTADTNVNDHNLPHFQILNKGITISVPINELKI 1237
Db 3 TPTPTSAAP---TPTPSAGGSLVQYRAADTNAGDNQKLPFRIVNRGTSSVPLSELT 59
QY 1238 RYYTIDGROTFNCYAVLSCKLNGKLVKMDKAAATGADYLYEVSFNSDAGVLASGG 1297
Db 60 RYWTYVGDGKQVFNCDWAQVGCNSVGRSFVKLSTGRTGADYIEITFTSGAGSLAAGS 119
QY 1298 TGGIOTRIHKADWSNYESDDYSYGTOTGFDHHTKATLYHNGVLVWGTTE 1347
Db 120 SGDIQVRINKNDWTNYEANDYSYDPTKTSFADNWRVTLYRNGQLVWVE 169

RESULT 13
Q9AAG9 PRELIMINARY; PRT; 209 AA.
AC Q9AAG9
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLYCOSYL HYDROLASE 5 (FRAGMENT).
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOK7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1."
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL; AF078039; AAK06389.1;
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 209 AA; 23217 MW; 4B283F2236D491DE CRC64;

Query Match 7.3%; Score 521.5; DB 2; Length 209;
Best Local Similarity 46.7%; Pred. No. 2.3e-19;
Matches 98; Conservative 36; Mismatches 63; Indels 13; Gaps 2;

QY 353 DPTYK-----EDSMIAQWNSSEFLPLRLKQSDYKYPGKTLALTEYSYGGENDI 402
Db 1 DPTYKTKQGOITAGENSINQWFEYLPPLNPKADIDKYPGTKLALTEEDYGGKDI 60
QY 403 SGGIAMADVGLIGKNDVYMANVWKLKDGANNVSAAYKLYRNYDGKSTFGDISVHAQT 462
Db 61 SGGIALADVGLIGFYGVYMAARW---GDSGYAQAAYNIYLYNDGKSGRYGTCVSAET 117
QY 463 SDIVNSSVHVSVDASYSKELHVLVNMKSMDSAFDAQFDLSETTGYSGKVGFDKNSQI 522
Db 118 TDVENNPVITASLEGDDSTVHIILINRNRYDKLKAIEKMNNRVTYGTGEIYGFDTSSQI 177
QY 523 KEAAPITQISGNRFTYTPPLTAYHIVLTA 552
Db 178 RKMVLSLNTQNTIITEVPLNLTIVYHIVLTS 207

RESULT 14
Q9Z4I1

```

ID Q92411 PRELIMINARY; PRT; 997 AA.
AC 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE CELLULOSE PRECURSOR (EC 3.2.1.4).
GN CELB.
OS Bacillus sp. Bp-23.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=89769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-23;
RX MEDLINE=21129642; PubMed=11234960;
RA Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., Diaz P.;
RT "Molecular cloning and characterization of a multidomain endoglucanase
RT from Paenibacillus sp BP-23: evaluation of its performance in pulp
RT refining.";
RL Appl. Microbiol. Biotechnol. 55:61-68(2001).
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL: AJ133614; CAB38941.1;
DR HSP: P26221; 1TF4.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR000566; Lipocln_cytfabp.
DR Pfam: PF00942; CBD_3; 2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR Prodom; PD001947; CBD_3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 34
FT SIGNAL 1 34
SQ SEQUENCE 997 AA; 106927 MW; F20CB119D6410629 CRC64;

Query Match
Best Local Similarity 6.7%; Score 477; DB 2; Length 997;
Matches 257; Conservative 156; Mismatches 427; Indels 328; Gaps 49;

QY 258 DAGAEIFG-----PVLVYGFQ-----AYKD-LQTPADNNSVKGNYSWFVDY 296
DB 79 DVGVDLTGGWDYDAGDHVHFGFPMASATMLAWSVVEYADYEQAGQLEIKDNRWATDY 138
QY 297 YLDQRLSSQAEKRLLDVDFVHWPYEMGGGIRITNEVGNDETKKARMQAPRTLWDPTY 356
DB 139 FM-----KAHTKPNELWG-----QVGAGNTDHA-----WGGP-- 165
QY 357 KEDSWIAQWNEFLPLRLKQSDVKYYPGKTLALTEYSYGENDISGGITAMADVILG 416
DB 166 ---AEVQMNRR-----PSFK---IDACSPGDLAA-----ETAALAAASSIV-FAD 204
QY 417 KNDVTMANY-----KLKDGANNY-----VSAAYKLYRNDYDKSSTFGDISVHAQT 462
DB 205 SDPVYSAKLLQHAKEYNFADTYRGKYTDCITDAAAFYNSTGYEDELAWGG----- 256
QY 463 SDIVNSSVHASVTASAYKELHLVWNKMSDAFDAQFDLSGETTYGSGKVGDFK----- 517
DB 257 -----AWLYLATDNAY-----LSKAL-SAADRWSTSGGSANWPYTWQGWDSKHYGA 303
QY 518 -----NSSOIKEAAPITQISGRFTYTPVPLTAYHIVLTAGNDTPVENPESFALRAE 569
DB 304 QILLARITNSLMPEATKFTOSTERNLDYTWGTNGGRVKYTPGGLAWDQWGLRYAAN 363
QY 570 AGDGKSIYLDASSGVVGVSVORATYENGPFVAASNLV-----ETSYTDNTVNTGTSY 623
DB 364 AAFISFVYSD-----WVSDPVKKSRYQN--FATSQINILGDNPRQSSVYGVYGVNSQPH 417
QY 624 YKITAKTGTGTS-----ASNLKAVPRAPVDGP---DRYEAEDGLTKGVVSESGTGFSGTG 677

Db 418 HHRTAHSSWMNEDIPANHRHILYCAMVGGPNASDQYTDIG-----D 460
QY 678 YVTN-----FHNAGDSLMTWTQAPTAGLYNLITGYRSPHDKRTNFSLNGKASGELVWKT 733
Db 461 YVSNEVATDYNAGF-----TGALAKNNLLYQNHQ-----PL 492
QY 734 ADF--KETSGGKVLNAGANTIGETGMWYDIDYKLEPAADPPHAPHAVTKTLINPATV 791
Db 493 ANFPAPVKGVDEYFVEAAVRSSG-----SNYT- 519
QY 792 EAKALMNYLDVQYCKNMLSGOEMPEIDWLQANVKKPAITA-ALDLID-YSPRAEHGLS 849
Db 520 EIRALL-----NNRSG-----W-PARGMDLSFKYFDLSEVYAAGRTVSDVQ 561
QY 850 STETEKAIEMWDKOGIVTFFAWHNNAKGLIDTOGKEWWRGFYADSTFDIEYAMNIPSE 909
Db 562 VTVS-----SSEGATVSQPVVDAAKRI-----YAITANF--SNTKIYPGGE 601
QY 910 DYKLLIRIDIDVIAGOLAKKLODAKVPVLPRLPHEAEGKWFWMGAKGP-----EPVKKL 961
Db 602 -----GNYRK-----EVQFR-ITGPQGAWN--PANDPSYQNLTTGCPVKSN 639
QY 962 YILMDRLTNVHKLNLLIWMVNSVAPDWPGEVDILTSPSYPOAGDYSPQIAKYEDLV 1021
Db 640 YIPYDAGVKVSGQE-----FVPTPVAVPAAPAGVQAVAGNSQVA-----L 680
QY 1022 TLGKDKKLVCHEKRRKTYPPGSGDEGVSSPLELVRYVMWGFLERRQTKQSLEHLKKVYNHP 1081
Db 681 NWSASAGAVSYTVKRAEVSGGPVTTVAAGVNGLTY-----TNTGLTNGKTYVVV 730
QY 1082 NVI-TLEKLTNLKTYGITE-QPSVPCSFLLNAGETAKVKLSWTASANAASVEYRSTV 1139
Db 731 TAVNAGESPASVQVSGMPOAATVPGAVTLSTAGNNQNLMSWTASGASSYTVQRA-V 789
QY 1140 ENGAFATVASDVYCGSSYTDFAVTADTYYQVVAKNADAGQTVTSASAAPKADTQOPTTG 1199
Db 790 AGTYTVDVATGLAVLNVDITLNGTISYRVAANVNSGOTLSNVVNLTPSGPPVSTGT- 848
QY 1200 LVQYRTADTNVNDNHLNPHQILNKGITISVPINELKIRYYITDGDREOTFNCDYAVLS 1259
Db 849 LEVQYRSGSGSGNSNAVTPQFNKNTGTQALDLSVKIRYFTKDTGTELSFWCDYAQVG 908
QY 1260 CSKLNGLKVKMDKAATGADYILEVSEFNSDAGVLASGSGTGIQTRHKADMSWNESDDY 1319
Db 909 SANVQGMFVAVNPAKGTADTYVEISTSGAGSLAAGTGVITQTRFSKNWNSAFDQSDNY 968
QY 1320 SYKGTQTSFDDHTKATLYHNGVVLWGTG 1347
Db 969 SYDASKTAPAAWKNVYATQNGTQVWGLE 996

RESULT 15
Q9X3P5
ID Q9X3P5 PRELIMINARY; PRT; 1770 AA.
AC Q9X3P5;
DT 01-NOV-1999 (TREMREL. 12, Created)
DT 01-NOV-2001 (TREMREL. 12, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE XYN.
GN XYN.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT thermophile Caldicellulosiruptor isolate Tok7B.1";
```


Db 202 CTGACGGTTCACAACTGGGAGAACCAACGCATCCAGCGCCGGAAGGACTGGCTTCATTAC 261
Qy 313 agcgataacitttttatcaacaatgggcttgacaaaaagccgaatgtgaaaaagccggga 372
Db 262 AGCGATGATTTCTCTCGGCAACGGTGGTTCAGACACCGACTGCGCAACAGCCGGG 321
Qy 373 gcaatgcagactctgttcacatgcctgaagctggcgcttattcttttagtcaag 432
Db 322 GCGGTTTACCGCTTTTCAGATAATCTTTGGAGAACTGGAGCTTACCTCCATTGAACG 381
Qy 433 ctgcgatggcgggttatgtggcaagatggaaaaggaagtgtgcaggaagacgaacag 492
Db 382 CTGCAAAATGGCGGTTATGTCTCCGGGATAAGAACCGTCCAGTTGACGAGAGTGAGACG 441
Qy 493 gctcttcgcctgttggaatcaggtcgtaaatgcacaaatgcgcgtttccaaactacag 552
Db 442 GCTCCGTCACCGCTTGGGATAGTGTGATTTGCCAAAATGCGCGTTCTCCCTTCAG 501
Qy 553 cctgatcctgaatgacatcaggttatgtcgatgaattctgcaatttttttagtgaagaaag 612
Db 502 CCTGATCTGAACGACGACGCAAGTGTATGATGAAGAAGTTAACTTCTCTGGTCAACCGG 561
Qy 613 tacggcgctgtctcaacaaaggcggtgtgaaaggatgatgcctgcacaaatgaacccgct 672
Db 562 TATGGAACCGCTTCAACGCTCAACGGGATCAAAAGCGTATTCGCTGGATAACGAGCGCGG 621
Qy 673 ctctgtgcgatcagcgcgcgcattcatgttggaagaaagtcnntnnnnnnnnnnnnnn 732
Db 622 CTGTGGTCTGAGCGCATCCAGGATTCATCCGGAGCAGTTACAGCGGCGAGACTCGTC 681
Qy 733 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 792
Db 682 GCTAAGAGCATCGACTTGTCAAGCGCGGTGAAGAACGCTCGATCCGATCGCGAAATATTC 741
Qy 793 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 846
Db 742 GGTCTGCGCTTACCGTTTTCGGCGCATATTTGTCTGCGAGCGCACCGGATTCGCGG 801
Qy 847 tctgtaaaagcaactagctgttgcgttgactattcactgatcaaatgcgcctcaac 906
Db 802 AGTTGACAGCACTACAGCTGTTTATCGATTAATCTGATCATGATGAAGAATGCT 861
Qy 907 tgcgaagccgaaggaagattgtcgatgtattctgatgtgcactgtgtatcccgagcg 966
Db 862 CATACGAGAACCGCAAAAGATTCCTCGATGCTGTGACGCTCCTGATTCGGAAGCA 921
Qy 967 atggcgaggagcatacgaattacaaatga---ggtagcaaatgacgaagaaagacc 1023
Db 922 CAGGCGGAGGCGCAAGTCTTTTGGCGGGCGGCAATTCGATACGCAAGAGCT 981
Qy 1024 agaatgcagcgctcgttactttgtggaccgcctcacaaggaagatagctggatcgt 1083
Db 982 CGCGTACAGCGCCCAAGATCCTATGGATGCGGCTTACCAGGAGACACTGGATCGCG 1041
Qy 1084 caatggaaacgcgacttcttgccttactgcctcgatgaagcagctgggtggaagat 1143
Db 1042 ACATGGTTTTCAGCTACTTGCCTTAAATCCGAAGCTGAATCTTCGATTACAGACGTAT 1101
Qy 1144 taccgggaacaaagctgtgttgacacagatagctacggcggaagaaatgatattcc 1203
Db 1102 TATCCGGGTACGAAGCTGGCGATCACAGATTCAGTACGCGGAGACAAATCATTTTG 1161
Qy 1204 ggcggtattgtatgacgatgtgtggtgcatcttgggcaaaacagctttttatggcg 1263
Db 1162 GGAGCATAGCTACCGCGGACGCTCGGCATTTTGGAAAATATGCGGTTTATGCGCG 1221
Qy 1264 aactattggaagttaaagatgtgtcccaacaactacgttttagcgccgttacaagctttac 1323
Db 1222 AATTACTGGCAG---ACGGAGGACAAATACCGATTTATACACGCGTGTCTTAAAGCTGAT 1278
Qy 1324 cgaattatgcgaaaaaacgacttcttcggcgatcatcgcttaatgcgcaaacgctg 1383

Db 1279 CGCAACTACGACGCAATAAATTCGGGGTTCCGCTCGATCAAAAGTGGACGCCGCTACGTCC 1338
Qy 1384 gatattgtaatagctcggtgcgtcttccgttaacggatgcacccatacaaaagaactgcac 1443
Db 1339 GATACGGAGAACAGCTCGGTATACGCTTCGGTAACTGACGAGGAGAAATTCGGAACCTCAC 1398
Qy 1444 ctcatgtcatgaataaaagcatggacagcgcatcgcagcccaaatctogattotttccggc 1503
Db 1399 CTGATCGTCTCAATAAATAATTCGACGATCGATCAACGCTTACTTTCAGCTGTCTGGT 1458
Qy 1504 gagaagacttacagttccggtaaaatatgggcttcgataaaaaatagctcgcaaatgaag 1563
Db 1459 GATAAAACCTACACATCCGGGAGATGCGGCTTCGACCAACCCGGATCCGACATTACG 1518
Qy 1564 cagtagcgcaatcacgcaaatcttcaggcaacgcgtttacctatacacgtacacaccttg 1623
Db 1519 GAACAAGCAGCTTAAACGAATATTAAACAATCAATTCACGTATACGCTTCTCCATTG 1578
Qy 1624 acggtttatcaatcggtgtgactgccgacaatgatcacacctgt 1667
Db 1579 TCGGCTTACACATGTTCTGAAAGCGGATAGCACCGAACCGGT 1622

RESULT 4
I73663
LOCUS I73663 1854 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 28 from patent US 5686593.
ACCESSION I73663
VERSION I73663.1 GI:3009804
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Woldike, H. Fabricius., Hagen, F., Hjort, C. Mailand, and Hastrup, S.
TITLE Enzyme capable of degrading cellulose or hemicellulose
JOURNAL Patent: US 5686593-A 28 11-NOV-1997;
FEATURES
Location/Qualifiers
1..1854
Source
BASE COUNT 452 a 506 c 503 g 393 t
ORIGIN

Query Match 26.6%; Score 569.8; DB 6; Length 1854;
Best Local Similarity 59.3%; Pred. No. 1.2e-158;
Matches 916; Conservative 0; Mismatches 616; Indels 12; Gaps 3;

Qy 133 gttaccattaaagtcgatacatccaaagatcgtaaagcctattagctcttataatacaggt 192
Db 82 GTCACCTTTCACGATTAATACGAGCTCGGAACGTCGAGCGATCAGCCCCCAATATTTACGGA 141
Qy 193 acgaatcagatttggcagcgcatgaaaatctgctgcagacgactgtgtgccaatcga 252
Db 142 ACAATCAGATCTGAGCGGAGGAGAACTGTCATCCGCGAGGCTCGGAGGCAACCGG 201
Qy 253 atgaccggatacaactgggaaaaataatagtccaaatgggggaagcgtattggcagcaatct 312
Db 202 CTGACGGTTTACAACCTGGGAGAACACGATCCAGCGCGGAAGGAGCTGGCTTCATTAC 261
Qy 313 acgataaactttttatgcaaatggtgcctgacaaaaagccgaatgtgaaaaagccggga 372
Db 262 AGCGATGATTTCTCTCGGCAACCGTGGTGTTCAGACACCGACTGCGCAACAGCCGGG 321
Qy 373 gcaatgcagactctgttcacatgcctgaagctggcgcttattcttttagtcaag 432
Db 322 GCGGTTTACCGCTTTTCACGATAAATCTTTGGAGAACTGGAGCTTACTCCATTGAACG 381
Qy 433 ctgcgatggcgggttatgtggcaagatggaaaaggaagtgtgcaggaagacgaacag 492
Db 382 CTGCAAAATGGCGGTTATGTCTCCGGGATAAGAACCGTCCAGTTGACGAGAGTGAGACG 441
Qy 493 gctcttcgcctgttggaatcaggtcgtaaatgcacaaatgcgcgtttccaaactacag 552

[illegible]

```

/transl_table=11
/product="endoglucanase"
/protein_id="BAAL2070.1"
/db_xref="GI:1663519"
/translation="MAKRRLLSLVLAIMFTVMVPPQISASAEVAPBGRKLLDVLQIF
KDSPPVWGSQMGLETFIGDPLVDVTYVNGPLRLNVQTVQSGWVLSLLTRG
WTHDLSQVGVNGVLEFDIKKEGEDFVIGFRDKVYERVYGLDVTYVLSNYVTYV
TDWHRVKTPLRLMKINNGPDPSSVTCVLFKRYADFTVWFSDIKITSEDNEKSAPA
IKVNLQGFIPAEKXVYALVTPGABELAVSEGDEFVINAADNSVAYTKGLTLVTEYEP
DSGKILKADSDLTVPCKYIYISIEGLDNPKEFEGEIGYGLVVDAAFYFORQGI
ELEPEPQVPRKDPQDAYAVFASGKDPIDITKGWDAGDFKYVNAAGATGVSOL
FWYEMFSPQVDFQFNIPESNGVPDILDEARWEMLKMODKESGVPYPRVQSDN
DENKLSRIIRQNGCTDDTACAGILAHAYLYVKDIDDPQAOECLDAAINAKMFELEK
NPNENVSPGPNVYDSDGRLWAAASLYRATGEVYHYTFKQNKYFAKFPSPRAY
AHTGDWLTAFSLYKAENKDOEVDWIDTEFGINLENTLRYENPNKNAIPVNG
FWGINKVMVPMDAALIGSQLGKYSRIEKLFGSLNLLGTLNLPNRPFSVSGEDS
KGVFNSIYNTDQKQIPGKPGPEPVPNTPPEAKVVDIRIDTSABRKPISPYIYGS
NSALVFQAFQSEVNPGPEPVPNTPPEAKVVDIRIDTSABRKPISPYIYGS
NOELDATVAKRFQGNRTTYNMENFNSNAGSDWLSYSDTYLLEDGVPKGENSTPAS
VVTTFHDKALSKNVPYTLITLOAGYVSADNGPVSQOETAPSRWKEVFKGAPPS
LTPQTEDVYVMDEKVNLYKNYGNASTPTGKIGYSIDNEPALMHHPRHDPNVA
KELIEKSVLSKVKVDPYAEIFGPALYGPAYETLQSDAPDMWTEGEGYRWFIDYIL
DKMKASDEBEKRLDVLVDHWPPEARGGERICFGADPRNIETNRKARLQAPRLDP
TYIEDSWIQWKKDFLPLNLDLSIEKYPTGLAITEYDYGNGHITGIGIAADVL
GIFKYGVLATFWGDSNNYTEAGINLYNYDKGKFGDTSVKCKTSIDIEVSSAVA
SIVGEDSKLHILLLKNYDQPTTFNFSIDSSKNYTIIGNVWAFDRGSSNLTORTPLVN
IKDNTFTYVTPALFACHIVLEAEPVYGLNDKSNVAVDILMKRYILIGIDINIL
TAADIYFGVNSSDYNIKRYLLKALIEDIPYVPEKQAPRAIFTFSPDPVTENVVF
NASNSIDEGTIIAYWDFGDEGTSTPTTYTYKKNPGTYKVKLIVTDNOGASSF
TATIKYSATGDSNKFNFDGTGGFTSGTNGATVNTYKAFKRGELKWTYSE
GEGFAELKLDGGTIVVPGTMTFRINIPSGAPIAAIQYIMPHPTPQWSEVLWNSTWK
YTMVKTDWNEITLITLPEVDPTWPOQMGIVQVOTIDBGEFTIYVDAIDW"
terminator
gene
5453..5491
CDS
5963..6465
/gene="celk"
5963..6465
/gene="celk"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="BAA20861.1"
/db_xref="GI:4432959"
/translation="MVKKTCFLVLTAMLFAFILPOGIVSAAGSYNYAELQKAIYFY
EQQAGPLPBNRWKMDATMDEVLGWYDAGDHVKFNLPMAYSAAMLGNALYEG
DTEASQORLHLENLAFALDYLVACDRGDSVYVYQIGDGAADHKWGSAEVIEKETR
PYFYGKDP"
BASE COUNT 2044 a 1182 c 1455 g 1784 t
ORIGIN

```

```

Query Match 20.0%; Score 428.2; DB 1; Length 6465;
Best Local Similarity 54.4%; Pred. No. 3.1e-116;
Matches 852; Conservative 0; Mismatches 687; Indels 27; Gaps 5;

QY 93 accggtcccaacttagttagcagtcgaacggaacggttaccattaaagtcgatac 152
|||||
DB 2886 ACCGGGAGTAGTCTCCGAATCTACAGAACCTGCAAAAGTGGTGCATCAGGATAGATAC 2945

QY 153 atccaaggatcgtaagcctattagcttatataatcaggtacgaatcaggtattgacgg 212
|||||
DB 2946 TTCTGCTGAAGAAGCAATACGCCGTATATATACGGAAGCAATCAGGAACCTT----- 3000

QY 213 cgtatgaatactggtgccagacgacttggtggcaatcggaatcgagatacaactggga 272
|||||
DB 3001 -GATGCAACAGTACTGCAAGAGGTTTCGGCGGAACAGAACTACAGGATACAACTGGGA 3059

QY 273 aaataaatatcgtaacggaacggttggcagcaatcgtacgatacaactttttatgcaa 332
|||||
DB 3060 AAACAACCTTCTCAAAATCGAGGAAGTACTGGCTGCATTACAGTGATACATACCTTTTGA 3119

QY 333 caatgggtgctgacaaaaacggaatgtaaaagcgggagcagtagtcgacttcgattca 392
|||||
DB 3120 GGACGGGGAGTTCCTTAAGGAGAGTGAGTACACCTGCTTCTGTAGTTACCACGTTCCA 3179

```

```

QY 393 tgatcaactcgct---gaagctgggcgcttattttagtcacgctgcgagtcgagcgtta 449
|||||
DB 3180 TGCAAGGCACCTTAGCAAAATGTTCTTACACACTTATCACCTTCAGGCAGAGGTTA 3239

QY 450 tgtggcaagagatggaacggaagtgctgaggaagcgaagcgaagcgtctcttcgctggtg 509
|||||
DB 3240 TGTTCGCGACGACGAACGACCGGTTTCCAGGAAGAAATGCAACCGTCTTCAAGATG 3299

QY 510 gaatcaggtcgtaaatcccaaaatcgccgttcccaactacagcgtgatatgtaatac 569
|||||
DB 3300 GAAGCAAGTAAAGTTTGAAAGGAGGACACCTTTCTCACTTACACCGGACACAGAGATGA 3359

QY 570 tcaggtatcgcgatgaattcgcataatttttagtgaataagtcagcgctgcttcaac 629
|||||
DB 3360 TTATGTTTACATGATGAGTTTGTAAACTATCTTGTAAACAAATACGGAATGCATCCAC 3419

QY 630 aaagcggtgtgaaagatatacgctcgcaaatgaacccgctctctgctcgatacga 689
|||||
DB 3420 ACCTACAGGAATAAAGGTTTATTCAATAGATAACGACCGCGCATTTGGAGTCACTCA 3479

QY 690 tcgcgcattcattggtgaaaaagtcmmnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 749
|||||
DB 3480 TCCGAGAAATTCATCCGACATGTAAGTGCACAAAGACTTATTGAAAAATCTGTAGCTCT 3539

QY 750 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 809
DB 3540 TTCCAAGCGGTTAAAAGGTAGATCCATATGCAGAAATATTTCGGACCTCTTTGTACGG 3599

QY 810 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 863
DB 3600 ATTTGCGCATATAGACACATTCAGTCAGCTCTCTGACTGGGAACTGAAGGAGAGGATA 3659

QY 864 tagctggttcgtagctattaccctggtgacaaatgcgctcaactcgaagcgaagcga 923
|||||
DB 3660 CAGTGGTATTATAGATTATTACCTCGATGAAGTGAAGAGGCTTCTGATGAAGAGGAAA 3719

QY 924 gaatgctgtagctgattcagctgtagctgtagctgtagctgtagctgtagctgtagctg 977
|||||
DB 3720 GAGCTTTTGGAGCTTGTGACATACCTGTTATTCGGAAGCCAGCGGGCGGTGAAAG 3779

QY 978 catacaattacaaatgagtaggcaatgacgaacgaagaaagcgaagtcgagcgcc 1037
|||||
DB 3780 AATATGCTTTGGAGCGCATCAAGAATATTGAGACAAACAAAGCAAGATTTCAGCGGCC 3839

QY 1038 tcgtaacttbtggtggacccacacacacacacacacacacacacacacacacacac 1097
|||||
DB 3840 CAGAACATTGTGGGATCCTACATATATTGAAGACAGCTGGATAGGACAATGGAAGAAGA 3899

QY 1098 attctgcttttactgctcgttaggaagcagtcggtggaacagattaccggggaaccaa 1157
|||||
DB 3900 TTTCTCCCGATATTACCTAAATCTTTTGGATTCCATTGAAAAATATTATCCGGGACGAA 3959

QY 1158 gctggtttgaccgagatagctacgcgcggaataatgatattccgcggtatgctat 1217
|||||
DB 3960 GCTTGCTATAACTGAATATGACTATGCGCGGAGAAATCATATTACAGCGGTATTGCTCA 4019

QY 1218 gaccggtgctggtggcatttgggcaaacacacacacacacacacacacacacacacac 1277
|||||
DB 4020 AGCCGATGTTCTTGGTATATTCCGTAATATACGTTGTTTACCTTGAACATTTCTGGG--- 4075

QY 1278 aaaggtggtgcaacaaactagcttagcgcccttacaagctttaccgcaatgaacgg 1337
|||||
DB 4076 --GAGATCAAGCAATACTATACCTAGCGCGGTATTAACCTTTTATACCACTACGAGG 4133

QY 1338 aaaaaacgctacttttcgcgatatacagcgttaatgcgaacacgctcggtattgtaataag 1397
|||||
DB 4134 CAAAGCGGCATATTGGAGATACATCCGTAAATGTGAAACGTCGACATAGAAGTAAG 4193

QY 1398 ctccgtgcatgctccgtaacggtatgcatcctacaaaagaaacacacacacacacacac 1457
|||||
DB 4194 CTCTGCTTATGATCCATTGTGGTGAAGATGACAGCAAACTCCATATATCTTTTGA 4253

```


KEYWORDS

SOURCE Ruminococcus flavefaciens.
 ORGANISM Ruminococcus flavefaciens
 Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 Ruminococcus.

REFERENCE

1 (bases 1 to 2492)
 Vercoe, P.E., Finks, J.L. and White, B.A.
 DNA sequence and transcriptional characterization of a
 beta-glucanase gene (celB) from Ruminococcus flavefaciens FD-1
 Can. J. Microbiol. 41 (10), 869-876 (1995)

JOURNAL

MEDLINE
 96025594

REFERENCE

2 (bases 1 to 2492)
 White, B.A.

AUTHORS

Submitted
 Direct Submission
 Submitted (11-APR-1994) Bryan A. White, Department of Animal
 Sciences, University of Illinois at Urbana-Champaign, 1207 West
 Gregory Drive, Urbana, IL 61801, USA

On Mar 29, 1995 this sequence version replaced gi:475751.

COMMENT

Location/Qualifiers

1..2492

/organism="Ruminococcus flavefaciens"

/strain="FD-1"

/db_xref="taxon:1265"

/clone="celB"

550..2448

/gene="celB"

550..2448

/gene="celB"

/function="endoglucanase"

/codon_start=1

/transl_table=11

/product="celB"

/protein_id="AAA95959.1"

/db_xref="GI:736357"

/translation="MRKSTAFLLTALAMAGCTLSAPGLVLPETKVNAGGFDNMKVDL

KGERKEISPLIYGVNQYTDLKSVKTTAVRQGNRTATYNNWNNASNAGSDKWHSSDN

NLSDSNPAEVRQLSKREAKYGVDMYTLQWAGYVSADKQGTVEDEVPASRWNE

VKFTGPAFDEPLDGVVYMEYVNIINKIGDSQSPGIGYSLDNEPVLWNDTH

PRVHPETVTEELGKNSIELAKVKPLDRAEFGPALYGYAFDLDDDEQHTESGD

VKSNNYHWLDCYLDQMKKABEEGRFLRDVLDIHYSESARTGAEDRVQSVRTLYE

EGFSENSWIGOWCMONVPIILPTIKSIDTYPGTKLAISEYNEKFGEDTGTQAQAE

LGCFADGVVLTALWGEPEFLISINLYTNVKGKGCFTGLIPASTEDVSKSTYAA

VNDKNSVPELPAEAAVMVVDSDAEDGEKIYEERKVTETEEFDDPSMINKNIV

EIPITDPEQVSKIVINDVTSSAGSGWATRLCMHQRKGQGRGLDIQELQSPAGQ

GSVRYRKY"

BASE COUNT 725 a 572 c 621 g 574 t

ORIGIN

Query Match 6.5%; Score 139.2; DB 1; Length 2492;
 Best Local Similarity 44.8%; Pred. No. 6.7e-30;
 Matches 623; Conservative 0; Mismatches 672; Indels 96; Gaps 5;

QY 134 ttaccattaaatcgcatacctcaaggatcgaagcctattagctcttatatacagta 193

DB 662 TGAATATCAAGTAGACCTCAAGGCGAGCGCAAGAGAGATAAGCCCTTTATCTACGGCG 721

QY 194 cgaatcaggattggcaggc-----gatgaaatctgctgcccagacgacttggtgcca 247

DB 722 TTAACCATGATACACAGACCTCAAGAGCGTCAAGACTACTGCTAGCTCAGGCGCGTA 781

QY 248 atcgaatgaccggatcacactggggaataatgltccaatgcccgaagcagattggcagc 307

DB 782 ACCGTATGACTCTTACAACTGGGGAACAATGCTTTAAACGCAGAGTTCCGAGCTGGAAC 841

QY 308 aatcagcagataactttttatgcaaatggtggcctgacaaaagccgaatgtgaaagc 367

DB 842 ACAGTTCCGATACAACTCTCGGATTCAAATCTCTCTGCTGAGGT----- 887

QY 368 cgggagcagtgacgactctgttcattgatcaatgctgaagctggcgcttattcttag 427

DB 888 -----AGTTCCAGAGACTTTTCAAGGAAGTGCACAAAGTACGGCGCTGATTACAAAATGA 940

QY 428 tcacgctgcgatggccggttatgtggccaagatggaaacacggaagtgctgcgggaaagcg 487

DB 941 CAACTCTCCAGATGGCAGGCTACGTTTCAGTGTATGAAGCAGCGCACTGTCAAGGAAGATG 1000

QY 488 aacaggctcttcgctcgtttggaatcaggtcgtaaatgccaaaaatgcgcgttcccaac 547

DB 1001 AAGTCGCTCTCTCAAAGCGCTGGAGCAGGTCAAGTTCACAAAGGCGCGACCTTTTGGCTG 1060

QY 548 tacagcctgatcgaatgacaatcaggtatatcggtatcggtgaattcgtcaatttttagtga 607

DB 1061 ATGAGCGCTGATCGACAGCGCGCTGTTTATATGATGAGTACGTAAACTACATCATCA 1120

QY 608 aagaatcagcgctgctctcaacaaagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 667

DB 1121 ATAAGCTGGCGGATTCACAGTCACTACAGGTATCCAGGCTACAGGCTTGATTAACGAGC 1180

QY 668 cgcctctgtgtgcatacgcctcgcgcattcatgtgtgaaaaagtcnnnnnnnnnnnnnnnn 727

DB 1181 CTGTTCTCTGGAACGATACTACCCGAGAGTTCATCCCGAGCTGTAACTATCGAAGAGC 1240

QY 728 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 787

DB 1241 TTGGAACAAGTCCATCGAATCTGCTAAGGCTCTCAAGAAGCTTGATCTCTAAGGCTGAGA 1300

QY 788 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 847

DB 1301 TCTTCGACCTGCTCTATGGCTACACTGCTTTCGATCACCTCGACGATGACGAGCAGC 1360

QY 848 ctgta-----aaaggcaactatagctggttcgtggaacttacc 886

DB 1361 ATACAGAATCGGGAGAGCTAAAGTCCCAAGAACAACTATCAGTGTATCTGGACTGCTATC 1420

QY 887 tggatcaaatgctcctcaactgcgaagcggaagagattgctggatgtattcgatg 946

DB 1421 TCACCCAGATGAAGAGGCTTCTGAAGAAGAGGACCCGCTCTCCTCGATGTACTTTGACA 1480

QY 947 tgcactgggtatcccggaagcgatggggcgaggagacacgaataacgaatgaggtaggcaatg 1006

DB 1481 TCCACTACTACTCGGAATCTGCAAGAACAAGGCT----- 1512

QY 1007 acgaaacgaagaagccgaatgcgctgctgactttgtgggacccgacccatacgaag 1066

DB 1513 -----GCTGAGGACAGAGTTCAGTCTGCTACTCTCTATGAGGAAGGCTTCTCCG 1564

QY 1067 aagatagctggatcgctcaatggaaacagcgcatcttgccttactgctcgcattgaaagc 1126

DB 1565 AGAACAGCTGGATCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1624

QY 1127 agtcggtggaagattaccgcgggaacacgctgcttgcgagatagctacgagcg 1186

DB 1625 AGTCCATAGATACATATATCCCGGCACAAAGCTGGCTATCTCAGAGTATAACTTCAAGG 1684

QY 1187 gcaaaatgatattccgcggtattgctatgaccgatgtgctgggcatcttggggcaaaa 1246

DB 1685 GCGCGAGGATACTTCGGTACTATCGCTCAGCAGAGGACACTGGGATGCTTCGCTGATC 1744

QY 1247 acgagctttatatggcgaaactattggaagttaagaatgggtgccaacacacgattagcg 1306

DB 1745 AGGCGGTATATCGCAACTCTCTGG-----GGCGGTGAGCCATTCATCATCT 1792

QY 1307 ccgcttaacgcttaccgcaattgacggaacacgctactcttgcgcatatcagcg 1366

DB 1793 CAGGTATCAACCTCTACATTAACCTACAGCGCAAGGCGGATGCTTCGGTGTACTTCTCA 1852

QY 1367 ttaatgcgcaaacgctcggtattgttaatgctcgtgtgcaatgcttcgcgttaacgagtgcat 1426

DB 1853 TCCCTGCTTCTACTGAAGATGTATCAAAAGTCCAGCACATACGCTGCTGTGTAAACGAGCGG 1912

QY 1427 cctacaagaactgcacctcattgtcatgataaagaacatgagcagcagcattcgagccc 1486

DB 1913 ATGAGTCAAAAGGTAACCTGTCATGATCACCACCAAGAAATATGACAGAAAGCTTGAGAACGCTG 1972

QY 1487 aattcgatctt 1497
 Db 1973 TTATCGACCTT 1983

RESULT 12

RFL298117 2692 bp DNA linear BCT 10-OCT-2001
 LOCUS Ruminococcus flavefaciens 17 endB gene for cellulase.
 DEFINITION AJ298117
 VERSION AJ298117.1 GI:16040919
 KEYWORDS cellulase; endB gene.
 SOURCE Ruminococcus flavefaciens.
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Ruminococcus.
 REFERENCE 1 (bases 1 to 2692)
 AUTHORS Rincon,M.T.
 JOURNAL Thesis (2001) Department of Agriculture, University of Aberdeen,
 Aberdeen, United Kingdom
 REFERENCE 2 (bases 1 to 2692)
 AUTHORS Rincon,M.T., McCrae,S.I., Kirby,J., Scott,K.P. and Flint,H.J.
 TITLE EndB, a Multidomain Family 44 Cellulase from Ruminococcus flavefaciens 17, Binds to Cellulose via a Novel Cellulose-Binding Module and to Another R. flavefaciens Protein via a Dockerin Domain Appl. Environ. Microbiol. 67 (10), 4426-4431 (2001).
 REFERENCE 3 (bases 1 to 2692)
 AUTHORS Flint,H.J.
 JOURNAL Direct Submission
 MEDLINE Submitted (13-DEC-2000) Flint H.J., Gut Microbiology Group, Rowett
 AUTHORS Research Institute, Greenburn Road, Bucksburn, Aberdeen, AB21 9SB,
 JOURNAL UNITED KINGDOM

FEATURES

source
 1..2692
 /organism="Ruminococcus flavefaciens"
 /strain="17"
 /db_xref="taxon:1265"
 /country="United Kingdom"
 176..271
 /gene="endB"
 176..2602
 /gene="endB"
 /function="beta-(1,4)-glucanase"
 /codon_start=1
 /transl_table=11
 /evidence=experimental
 /product="cellulase"
 /protein_id="CAC83072.1"
 /db_xref="GI:16040920"
 /translation="MKKTAFAIAVCVSGTMTAPVNLPRATVNAAGGYDMNVVDL
 KGEKATSLPIYGVNOVTTDLRDVKTAVRQGNMAYNWEENASNAGSDWKHSDN
 NLSDDPACVOVLKQAAKYVNVYKLTLLQAGVYSADKNGPVSEAKRPSDRWNK
 VYLTKNAPFADTDLDTGVYMDIYVNIINKLGSQSAEIGIYSLDNPFLVNDTH
 SRMPDPTIEELGSKSEVEMAKVKLDKPAEYFPAIYGTAFDHLDDDDAHTWEE
 IKKANTHWIDCYLDHMHKASENGARLDVLDIHYSESARKGIEDRLQSVRTLYE
 PGFSENIOWCMENVPILPTQKSIDTYPGTKLGISEYNGFGDDASGTTAQAEF
 LGCYADGVVPSLWNGEPEFLSGIQLYTNVKGKCGFGLIPASTGDVSKSTYAA
 VNKADDSKVTVMTNKLKENENAVIDLNRADKSYKSAAYVAFVGDSEIRLIDILIK
 VKNKVTETLPASAAVVSDDQADAFDGLKTEETKTKTFVETDIESMTNKGFEV
 VPIEDAEHLKILINGVSSAGSNWATAGCAVMNVTAKDGSFTWYKSYNLPGLS
 KSSATFEDGIFTKETGEGADKYKEDLEATVAREKVELQKWDASKESSSEDKTEV
 KYSSIQVVEYAOGEAPKPVTTTTSSAVTTTTSVATTTSSTTAPDKDYYVGANDCG
 NVLDAILILMOSLANPNFKGLNSAEKHLTEKGRNADCCDPGSGILTNDDALSIRL
 LLHLIDKLPE"
 176..2602
 /gene="endB"
 272..2599
 /gene="endB"
 /product="EndB protein"

CDS

721 a 683 c 724 g 564 t
 gene
 mat_peptide
 BASE COUNT 721 a 683 c 724 g 564 t
 ORIGIN

Query Match 6.3%; Score 135.6; DB 1; Length 2692;
 Best Local Similarity 44.2%; Pred. No. 8e-29;
 Matches 456; Conservative 0; Mismatches 548; Indels 27; Gaps 3;

QY 430 acgctgcccagtgccggttatgtgccaagatggaacaggaaagtgtgcaggaaagcga 489
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 569 ACCCTCCAGCTTGAGGCTATGTTTCAGCTGATAAGAACGGCCCTGTTCCGAGCGGAA 628
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 490 caggctccctccgctcgttggaatcaggctgaaatgccaaaaatgcgcggttccaacta 549
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 629 AAGCTCCCTCTGACCGCTGGACAAGTGGTGTACAAAGATGCTCCCTTTGCCGAT 688
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 550 cagcctgatcgaatgaacatcaggtatcggtatcggaattcgtcaatttttagtgaaa 609
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 689 ACTCCGACCTTCACAGACGGAGTTGTCTACATGACGAGTACGTGAACATATATAAT 748
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 610 aagtcagcgctgcttcaacaaaggcggtgtgaaagatatcgctcgacaatgaaccc 669
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 749 AAGCTGGCGGATTCACAGTGGCTGAGGCTATCCAGGCTTACAGCCTTGACAATGAGCC 808
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 670 gctcctggtgcatacgcatccgcgcatcattcgtgaaagagtcnnnnnnnnnnnnnn 729
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 809 GTGCTCTGGAACGATACACAGCAGGATGCACCTTCATCCGGTTACTATCGAGGACTG 868
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 730 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 789
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 869 GGCAGCAGTCCGTTGAATGGCAAGGCTGTAAAGAGCTGGATCCTTAAGGCTGAGGTG 928
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 790 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 849
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 929 TTCGGCCCGCTCTCTACGGCTACACCGCCTTCGATCACCTTTGACGAGATGACGCTAT 988
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 850 gtaaaaggcaactagctggttcgtgactatttaccctggtacaaatgcgcctcaactcg 909
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 989 ACCGAGTGGGAGGAGATAAGAAAGGCCAATAATTACCATGGTATCTGGACTGCTACCTC 1048
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 910 caagccgaagcaagagatgctggtatgctgagtgcactgggtatcccgaaagcgtg 969
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1049 GATCATATGCACAAGGCTCCGAGGACACGGCGCAAGACTTCTTGA-----CGTTC 1100
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 970 ggcggaggcatacgaattacaatgagtaggcaatgacgaacgaagaagcgaatg 1029
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1101 TGGATATACATTATTTCGGAGTCACCAAGAAAGG-----CATCGAGGACAGCTC 1153
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1030 caggcgctcgtacttbtgtggaccgcacctacaagaagatagctggatcgctcaatgg 1089
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1154 CAGTCTGTGCGCACACTCTATGAGCCGGCTTCTCAGAAAAACAGCTGGATAGGTCACTGG 1213
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1090 aacagcgacttctgcttactgcctcgattgaagcagtcggtgacaaagtattaccgg 1149
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1214 TGGATGGAGAAGCTGGCGATATTCCTACCATACAGAAGTCCATCGATCTTACTATCTC 1273
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1150 ggaaccaagctggtcttgaccgagtagtagtaacgagcggaagaaatgatattccgcggt 1209
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1274 GGCACAAAGCTGGGCATATCCGAGTACAACTTCGGCGGGCGGCACCATGCTTCGGCACT 1333
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1210 attgctatgaccgatgtgctgggacttcttgggcaaaaagacgcttattatggtgcaactat 1269
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1334 ATCGCTCAGGCTCAGGCTCTGGGCTGCTATCGCGATCAGGGCGTGTACTTTGCTCCCTC 1393
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1270 tgaagttaagatggtgccaacactacgtttagcgcgcttacaagctttaccgcaat 1329
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1394 TGG-----GGCGGCGAGCCCTTTATTTCTCGGGTATCCAGCTTATATACAAAC 1441
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1330 tatgcgaaaaaacactacttttcggcgatatacagcggttaatgcgcaaacgctcgatatt 1389
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1442 TATGACGGCAAGGCGGCTGCTTCGGCGGATACCTTATCCCTGCATCTACCGGGGATGTT 1501
 QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1390 gttaatagctcgggtgactcttcgtaacgagatgcatcctacaagaagactgcactcatt 1449
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1502 TCCAAGTCCAGCACCTATCGGCTGTAAACGAAAGGACGATTCAAAGGCTACTGTTATG 1561


```
source
1..7162
/organism="Kluyvera ascorbata"
/strain="Cip82.95m"
/db_xref="taxon:51288"
1067..2272
/gene="aspat"
1067..2272
/cds
/transl_table=1
/product="putative aspartate amino transferase"
/protein_id="CAC03612.1"
/db_xref="GI:984377"
/translation="MADSSPERRFRTRIDRLPPYVFNITAEKMAARRRGEDIIDFSMG
NPDGATPHIVKELCTVAORPDTHGYSTSGIPRLRAISRFWQERYDVELDPEDEAI
VTIGSKGLAHLMLADLHDGDTLVLPNPSYPIHIYGAVIAGIAQVRSLVGVDFNFE
LERAIRESYPPKMMILGFPNPTAQCVLEFFKFKVALAKRKYDLVVLHDLAVIVF
DGMKAPSIQVPGARDVAVFEFTLSKSYNMGWRIGFMVGNKDTLVNALARIKSYHDYG
TFTPLQVAAIAALEGDOQCVDLIAQYKRRRDVLVKGHEAGWMDMPKASMYVWAKI
PEQAHMGSLEFAKKLQDAKVCVSPGIGFDYGDTHVRFALLENRRDIRQAIRGKA
MFRADGLLP"
2978..3853
/gene="kluA-1"
2978..3853
/gene="kluA-1"
/codon_start=1
/product="beta-lactamase class A"
/protein_id="CAB59824.1"
/db_xref="GI:6165409"
/translation="MTQSIIRSMUTVMATPLPLIFSSATLHAQNSVQOOLEAKKSS
GRLGVALLINTADNSOILYRADERFAMCSYKVMAAAVLVKQSESKHLNQRVEIKK
SDLVNYPILAEKHVNGTMTLAEIGAAALQISDNTAMNKLTIAHGGPKDVTAFARSIGD
ETFLDTEPTLTAIECPDRDITPLAMQATLKNLTGLKALAEATQRAQVTLWKGNT
TGSASIRAGLPKSVVGVGDKTSGSDYGTNDIAITWPNHAPLVLTVTFTPEQKAEER
RDVLAATAKIVTHGF"
4277..4924
/feature="ORF3"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAC03613.1"
/db_xref="GI:9843778"
/translation="MDKEGTMSLNNNFARVTCMSPCHCEVKNASLRVSCQECYCF
CTWPEASRYSLGIMEGLLKQOYHLYLGVYVDFYSILYRFTDKRDLDTFETLKL
IYLVCDRLPLANTYKHKHDIPLIYQDNLKSAGEKLRKFIYORDAFFHGLSL
TLEFDVLGALTAGENCNHLARGHRLDIRAVYAAKRAEKKMGADINTLRFSSHSS"
5144..7102
/feature="ORF4"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAC03614.1"
/db_xref="GI:9843779"
/translation="MMPKLERKPLALCITALLMSNFYCTTAHAETELATTLLYYVG
SGNSNGNSDFTNVDENVTISDWGTTSGLKGWALYMGENYKFNDFLVKTKGMAD
GIHTKNGGNIYENYNAITSGYSSDGINLRELNANYSRVTNNETNIAKTMGVRA
NASINNANAVATITLPGTTTETTSAGSVAGYVAGQDTEANSTAVKGAEIFLNG
DSTLTKVGRSAHAYARAAGTVSLGNVNISAMGDLANGLFASSSSAQQAAYLAKDA
VINVTGSAYAGAEALVKYDIDSGTEVAGYVLINGINIVADNDQGVSLHWADGSVF
TGITNSQLATSAANAGDIALQIDGNSLWMTKSDVLSLTLNNGCTVEWAGASAGNV
LTVAGNYOGNNGLLRLINTVLGDDNSVDKLVVBDTSGTTEAVTNAGSGDKTINGI
EVTHVDSQDGNFVQSGRIVAGSYEVALRGECTAFNHVYLNAAITPEPVERPTL
RPSGSSYLANAANFTYTRLDRLGETQYIARLGEQKVTSMWNNVGGHTRFKDS
SQGLSTNSRYIQLQGDIAQSTNGLDRLHGLMAGYNSQSNKTSKSDVSRHSGQV
TGSAGIYGTWAYNEADKTSYLDSWIH"
BASE COUNT 1800 a 1692 c 1856 g 1814 t
ORIGIN
```

Query Match

3.3%; Score 69.6; DB 1; Length 7162;

Best Local Similarity

72.6%; Pred. No. 5e-09;

```
Matches 90; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Qy 2018 ccgggtctcogtgactggttatgaactaattcccaatgaggggattccctgacga 2077
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7029 CCGGGAITTTATGGCACCTGGTATGCAAAACGAAGCGGATAAAACGGCAGTTATCTGGATA 7088
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2078 tgacatccactactgctgacactgcagcgcgagctccagcttttgcctcttagtg 2137
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 7089 GCTGATCCACTAGTGTGACCTGCAGGCGCGGAGCTCCAGCTTTTGTTCCTTTAGTG 7148
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 2138 aggg 2141
|||||
Db 7149 AGGG 7152

RESULT 15
AF322456/c
LOCUS
DEFINITION
AF322456.1 GI:11559861
VERSION
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 51381)
Zhuo,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
Gu,J.R.
Gene clone on human chromosome 17p13.3
Unpublished
2 (bases 1 to 51381)
Zhuo,X.T., He,M., Wan,D.F., Ye,Y., Qin,W.X., Huang,Y., Zuo,L. and
Gu,J.R.
Direct Submission
Submitted (21-NOV-2000) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai
200032, P. R. China
* NOTE: This is a 'working draft' sequence. It currently
* consists of 120 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 236: contig of 236 bp in length
* 237 gap of unknown length
* 435: contig of 199 bp in length
* 436 gap of unknown length
* 665: contig of 230 bp in length
* 666 gap of unknown length
* 879: contig of 214 bp in length
* 880 gap of unknown length
* 1121: contig of 242 bp in length
* 1122 gap of unknown length
* 1948: contig of 827 bp in length
* 1949 gap of unknown length
* 2629: contig of 680 bp in length
* 3230: contig of 601 bp in length
* 3893: contig of 664 bp in length
* 3894 gap of unknown length
* 4090: contig of 197 bp in length
* 4091 gap of unknown length
* 4272: contig of 182 bp in length
* 4273 gap of unknown length
* 4420: contig of 148 bp in length
* 4421 gap of unknown length
* 4633: contig of 213 bp in length
* gap of unknown length
```

* 4634	5219:	contig of 586 bp in length	* 21418	21679:	gap of unknown length
* 5220	5881:	gap of unknown length	* 21680	21888:	contig of 262 bp in length
* 5882	6118:	contig of 662 bp in length	* 21889	22072:	gap of unknown length
* 6119	6334:	gap of unknown length	* 22073	22788:	contig of 209 bp in length
* 6335	6484:	contig of 237 bp in length	* 22789	23448:	contig of 184 bp in length
* 6485	6623:	gap of unknown length	* 23449	23658:	gap of unknown length
* 6624	6861:	contig of 150 bp in length	* 23659	23869:	gap of unknown length
* 6862	7067:	gap of unknown length	* 23870	24449:	contig of 211 bp in length
* 7068	7836:	gap of unknown length	* 24450	25107:	gap of unknown length
* 7837	8541:	contig of 139 bp in length	* 25108	25316:	contig of 580 bp in length
* 8542	8731:	gap of unknown length	* 25317	25526:	gap of unknown length
* 8732	8875:	contig of 144 bp in length	* 25527	25902:	contig of 376 bp in length
* 8876	9094:	gap of unknown length	* 25903	26555:	gap of unknown length
* 9095	9684:	contig of 219 bp in length	* 26556	26764:	contig of 653 bp in length
* 9685	10350:	gap of unknown length	* 26765	26973:	contig of 209 bp in length
* 10351	10528:	gap of unknown length	* 26974	27587:	gap of unknown length
* 10529	11128:	contig of 178 bp in length	* 27588	27796:	contig of 614 bp in length
* 11129	11784:	gap of unknown length	* 27797	27997:	gap of unknown length
* 11785	11991:	contig of 600 bp in length	* 27998	28229:	contig of 209 bp in length
* 11992	12648:	gap of unknown length	* 28230	28451:	contig of 232 bp in length
* 12649	13323:	contig of 657 bp in length	* 28452	29087:	contig of 222 bp in length
* 13324	13607:	gap of unknown length	* 29088	29741:	contig of 636 bp in length
* 13608	13833:	gap of unknown length	* 29742	29980:	gap of unknown length
* 13834	14452:	contig of 226 bp in length	* 29981	30204:	contig of 239 bp in length
* 14453	15107:	gap of unknown length	* 30205	30725:	contig of 224 bp in length
* 15108	15743:	contig of 655 bp in length	* 30726	31377:	gap of unknown length
* 15744	16395:	gap of unknown length	* 31378	31587:	contig of 521 bp in length
* 16396	16614:	contig of 636 bp in length	* 31588	31792:	gap of unknown length
* 16615	16826:	gap of unknown length	* 31793	32403:	contig of 205 bp in length
* 16827	17628:	contig of 212 bp in length	* 32404	33056:	gap of unknown length
* 17629	18323:	gap of unknown length	* 33057	33266:	contig of 611 bp in length
* 18324	18532:	contig of 802 bp in length	* 33267	33471:	gap of unknown length
* 18533	18742:	gap of unknown length	* 33472	34137:	contig of 205 bp in length
* 18743	19453:	contig of 210 bp in length	* 34138	34789:	gap of unknown length
* 19454	20109:	gap of unknown length	* 35003:	contig of 666 bp in length	
* 20110	20724:	contig of 711 bp in length		gap of unknown length	
* 20725	21417:	contig of 656 bp in length		contig of 652 bp in length	
		gap of unknown length		gap of unknown length	
		contig of 615 bp in length			
		gap of unknown length			
		contig of 693 bp in length			
				3.0%; Score 63.2; DB 2; Length 51381;	

Query Match

c:\prog file\ winrar\

- Winrar - zip

- ~~30 days demo~~ ace

- Crack.exe. rar

cab

ark
c

175x2